Transformation-Invariant Learning of Optimal Individualized Decision Rules with Time-to-Event Outcomes

Yu Zhou, Lan Wang, Rui Song and Tuoyi Zhao

Abstract

In many important applications of precision medicine, the outcome of interest is time to an event (e.g., death, relapse of disease) and the primary goal is to identify the optimal individualized decision rule (IDR) to prolong survival time. Existing work in this area have been mostly focused on estimating the optimal IDR to maximize the restricted mean survival time in the population. We propose a new robust framework for estimating an optimal static or dynamic IDR with time-to-event outcomes based on an easy-to-interpret quantile criterion. The new method does not need to specify an outcome regression model and is robust for heavy-tailed distribution. The estimation problem corresponds to a nonregular M-estimation problem with both finite and infinite-dimensional nuisance parameters. Employing advanced empirical process techniques, we establish the statistical theory of the estimated parameter indexing the optimal IDR. Furthermore, we prove a novel result that the proposed approach can consistently estimate the optimal value function under mild conditions even when the optimal IDR is non-unique, which happens in the challenging setting of exceptional laws. We also propose a smoothed resampling procedure for inference. The proposed methods are implemented in the R-package `QTOCen`. We demonstrate the performance of the proposed new methods via extensive Monte Carlo studies and a real data application.

KEY WORDS: Exceptional laws, individualized decision rule, inference, precision medicine, robust method, time-to-event data.

[1]Yu Zhou is a machine learning engineer at Roku. Email: izhou@roku.com. Lan Wang is Professor, Department of Management Science, University of Miami. Email: lanwang@mbs.miami.edu. Rui Song is Professor, Department of Statistics, North Carolina State University. Email: rsong@ncsu.edu. Tuoyi Zhao is a Ph.D. student, Department of Management Science, University of Miami. Wang, Zhou and Zhao's research was supported by NSF FRGMS-1952373. Song's research was supported by NSF DMS-2113637.

1 Introduction

The problem of estimating the optimal individualized decision rule (IDR) has recently received substantial attention in precision medicine and other domains. A treatment can be a drug, a therapy, or any other actionable choice (or a sequence of such choices) such as a policy or program. The goal of optimal IDR estimation is to determine a decision rule that assigns a subject to one of the treatment options based on individual information available each decision point such that some functional of the potential outcome distribution is optimized.

For completely observed data, several successful approaches exist for estimating the optimal IDR, including Q-learning (Watkins and Dayan, 1992; Murphy, 2005b; Chakraborty et al., 2010; Song et al., 2015), A-learning (Robins et al., 2000; Murphy, 2003, 2005a; Moodie and Richardson, 2010), model-free or policy search methods (Robins and Rotnitzky, 2008; Orellana and Robins, 2010; Zhang et al., 2012; Zhao et al., 2012, 2015b), the interpretation-enhanced tree or list-based methods (Laber and Zhao, 2015; Cui et al., 2017; Zhu et al., 2017; Zhang et al., 2018) among others. See also the books of Chakraborty and Moodie (2013) and Kosorok and Moodie (2016) for a general introduction and additional references.

The focus of this paper is on estimating the optimal static (one-stage) or dynamic (multi-stage) IDR for time-to-event or survival data, where the outcome is possibly censored. Several new challenges arise when analyzing such data compared with the complete data case. Censoring occurs when an individual drops out from the study or the study ends before the subject experiences the event of interest. The distribution of survival time (e.g., time to death, onset of disease) is often highly skewed. The situation gets even more complicated for estimating the optimal dynamic IDR, where the data are collected longitudinally. Consider the setting where the treatment decisions for a patient are made at k pre-specified decision points. Each decision is allowed to depend on the patient's characteristics (e.g., gender, age) and treatment history (e.g., disease progression status and how the individual responds to previous treatments) up to that decision point. The patient may be censored at any stage of treatment. Direct application of existing complete-data techniques could result in severe bias, as demonstrated in the Monte Carlo studies in Section 5.

Several authors have recently investigated estimating the optimal IDR with survival data, see Goldberg and Kosorok (2012), Xu et al. (2016), Jiang et al. (2017a), Jiang et al. (2017b), Bai et al. (2017), Hager et al. (2018), Xu et al. (2016), Díaz et al. (2018), Simoneau et al. (2020), among others. For time-to-event outcomes, new criterion is needed to evaluate the effectiveness of an IDR. Of crucial importance is that such a criterion can be reliably estimated under censoring. The existing work have been mostly focused on maximizing the *restricted mean survival time* in the population.

We consider time-to-event outcomes and propose a new robust framework for estimating the optimal IDR using an alternative criterion based on the marginal quantile of the potential outcome distribution. The new optimality criterion is easy to interpret. Median survival time has already been popularly used clinically to evaluate the success of cancer treatment. It can be reliably estimated even under relatively heavy censoring (e.g., the censoring rate is more than 50% in the real data example of this paper). The resulted optimal IDR is invariant under monotone-transformation of the outcome. We develop robust estimation methods for both static and dynamic optimal IDRs. The robust approach circumvents the difficulty of specifying a reliable outcome regression model, especially for the dynamic setting which demands a sequence of generative regression models.

We consider estimating the optimal IDR in a class of candidate decision rules indexed by a finite-dimensional parameter. The estimation problem corresponds to a challenging non-regular M-estimation problem with both finite and infinite-dimensional nuisance parameters, due to the unknown censoring distribution. For the optimal static IDR, we rigorously establish the cube-root convergence rate for the estimated parameter by employing modern empirical processes techniques. We prove that its asymptotic distribution corresponds to the maximizer of a centered Gaussian process with a parabolic drift. It is worth emphasizing that the nonstandard asymptotics is due to the intrinsic nature of the decision problem, which relates to a sharp edge effect in the decision function. Due to the nature of nonstandard asymptotics, the problem is substantially harder than regular M-estimation problem with infinite-dimensional nuisance parameters.

Moreover, we establish a useful novel result that shows the optimal value can be consistently estimated under weak conditions for the challenging setting of exceptional laws.

Under exceptional laws, there exists a subgroup of patients for whom the treatment is neither beneficial nor harmful. In this case, the optimal IDR is non-unique, see for example, the discussions in Robins and Rotnitzky (2014) and Luedtke and van der Laan (2016).

Theoretically, our work complements existing results and significantly enhances the knowledge about optimal IDR estimation with survival data. The existing results have been mostly focused on prediction error bounds and have not studied the properties of the estimated parameter indexing the optimal IDR. Furthermore, to the best of our knowledge, existing work on optimal IDR estimation with survival outcomes assume non-exceptional laws and thus avoid the problem of optimal value estimation under exceptional laws.

The rest of the paper is organized as follows. Section 2 introduces the new framework and the robust method for estimating an optimal static IDR with survival data, with theoretical properties developed in Section 3. Section 4 presents the estimation method and the theory for the dynamic IDR setting. In Section 5, we report results from extensive Monte Carlo studies. In Section 6, we illustrate the application on the analysis of a breast cancer data set. The proposed methods are implemented in the R-package `QTOCen`. The regularity conditions are given in the Appendix. The technical derivations and additional numerical results are given in the online supplement.

2 Robust estimation for static optimal IDR

2.1 Preliminaries

We first consider the single-stage setting. Let $A \in \{0, 1\}$ denote the binary treatment, $\boldsymbol{X} \in \mathbb{R}^p$ denote the vector of covariates with support $\mathcal{X}$, and $T \in \mathbb{R}^+$ denote the time to the event of interest (or a transformation thereof). We often refer to T as *survival time*. Without loss of generality, we assume that a larger value of T indicates better treatment effect. The outcome T may not be observed due to censoring. Let $C \in \mathbb{R}^+$ denote the censoring variable and $\Delta = I(T \leq C)$. If the observation is censored (i.e., $\Delta = 0$), then we only observe C. Let $Y = \min\{T, C\}$ be the observed outcome. The observed data consist of $\{\boldsymbol{X}_i, A_i, Y_i, \Delta_i\}$, $i = 1, \ldots, n$, which are independent copies of $\big(\boldsymbol{X}, A, Y, \Delta\big)$.

To assess the treatment effect, we adopt the potential outcome framework (Neyman, 1923; Rubin, 1978) in causal inference. The i-th subject has two potential outcomes: $T_i^*\big(0\big)$ and $T_i^*\big(1\big)$, where $T_i^*(0)$ is the survival time had the subject received treatment 0 and $T_i^*(1)$ is defined similarly, $i = 1, ..., n$. In practice, a subject receives one and only one of the two possible treatments. Under the stable unit treatment value assumption (Rubin (1986)), we have $T_i = A_i T_i^*\big(1\big) + \big(1 - A_i\big)T_i^*\big(0\big)$. That is, the survival time of the ith subject is the potential survival time corresponding to the treatment that subject actually received. Furthermore, we assume the no unmeasured confounders assumption (Rosenbaum and Rubin, 1983) is satisfied, i.e., $\{T_i^*(0), T_i^*(1)\}$ are independent of A_i conditional on $\boldsymbol{X}_i$. This is a common assumption in causal inference and is automatically satisfied for a randomized trial. Mathematically, an IDR $d(\boldsymbol{X})$ is a mapping from the space of covariates $\mathcal{X}$ to the space of candidate treatments $\{0, 1\}$. The potential outcome associated with $d(\boldsymbol{X})$ is denoted by $T^*\big(d(\boldsymbol{X})\big)$. We have $T^*\big(d(\boldsymbol{X})\big) = T^*\big(1\big)d\big(\boldsymbol{X}\big) + T^*\big(0\big)\big[1 - d\big(\boldsymbol{X}\big)\big]$.

2.2 New optimality criterion for evaluating IDR with time-to-event outcome

Different from the completely observed data case, the mean of the outcome is usually difficult to estimate accurately at the presence of censoring. The most popular criterion in the existing literature for time-to-event data is the restricted mean survival time $\mathrm{E}\{\min\big(T^*\big(d(\boldsymbol{X})\big), L\big)\}$, where L is a user-supplied cut-off time. In this paper, we consider an alternative criterion for comparing IDRs for time-to-event outcomes based on the marginal quantile treatment effect. The new criterion enjoys three appealing properties: easy to interpret, robust to long-tailed survival distribution, and invariant to the monotone transformation of survival time. The marginal τ-th quantile ($0 < \tau < 1$) of the potential outcome $T^*\big(d(\boldsymbol{X})\big)$ is defined as

$$Q_\tau\big\{T^*\big(d(\boldsymbol{X})\big)\big\} = \inf\Big\{t \in \mathbb{R} : P\big\{T^*\big(d(\boldsymbol{X})\big) \le t\big\} \ge \tau\Big\},$$

where P denotes the marginal distribution of $T^*\big(d(\boldsymbol{X})\big)$. In IDR estimation problems, $Q_\tau\big\{T^*\big(d(\boldsymbol{X})\big)\big\}$ is referred to as the *value function*. We sometimes use the short-hand

notation $Q_\tau\{T^*(d)\}$. Given a class $\mathcal{D}$ of candidate IDRs, the optimal IDR is defined as

$$d_{opt}(T^*\big(d(\boldsymbol{X})\big)) = \arg\max_{d\in\mathcal{D}} Q_\tau\big\{T^*\big(d(\boldsymbol{X})\big)\big\}. \tag{1}$$

As an example, the above optimal IDR with $\tau = 0.5$ will lead to the maximal median of the potential survival time if each individual in the population follows the treatment recommended. In some other applications (e.g., distributing aid in a social welfare program), we may target improving the lower quantile by considering a smaller value of τ. For an arbitrary monotone transformation $h(\cdot)$, it is known that $h\big(Q_\tau\{h(T^*\big(d\big))\}\big) = Q_\tau\{h(T^*\big(d\big))\}$ (Koenker, 2005). Hence, the same decision will be reached no matter the analysis is based on the original survival time or its monotonic transformation.

For complete data, quantile criterion was studied in Wang et al. (2018) but their work is not applicable to survival data, which is also theoretically substantially harder with the presence of an infinite-dimensional nuisance parameter due to censoring. Wahed (2009) studied estimating the survival quantiles in two-stage randomization designs with fixed IDRs but had not investigated the more challenging problem of optimal IDR estimation.

In practice, $\mathcal{D}$ is usually chosen to be a class of IDRs indexed by a Euclidean parameter for interpretability. Same as Zhang et al. (2012) and others, we focus on the class of index rules $\mathcal{D} = \Big\{d_{\boldsymbol{\beta}}\big(\boldsymbol{X}\big) = \mathrm{I}\big(\boldsymbol{\beta}^T\boldsymbol{X} > 0\big) : \ \ |\beta_1| = 1,\ \widetilde{\boldsymbol{\beta}} \in \mathbb{B}\Big\}$, where $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T = (\beta_1, \widetilde{\boldsymbol{\beta}}^T)^T$, $\mathbb{B}$ is a compact subset of $\mathbb{R}^{p-1}$ and $\mathrm{I}(\cdot)$ denotes the indicator function. For identifiability, we assume there exists a continuous covariate whose coefficient has absolute value one. Without loss of generality, we assume $|\beta_1| = 1$. The population parameter $\boldsymbol{\beta}_0 = (\beta_{01}, ..., \beta_{0p})^T$ indexing the optimal IDR is

$$\boldsymbol{\beta}_0 = \arg\max_{\boldsymbol{\beta}\in\mathbb{B}^o} Q_\tau\{T^*\big(d(\boldsymbol{X})\big)\},$$

where $\mathbb{B}^o = \{\boldsymbol{\beta} \in \mathbb{R}^p : |\beta_1| = 1,\ \tilde{\boldsymbol{\beta}} \in \mathbb{B}\}$.

2.3 A robust estimation procedure

Based on the observations $\big\{\boldsymbol{X}_i, A_i, Y_i, \Delta_i\big\}$, $i = 1, \ldots, n$, our goal is to estimate the population parameter $\boldsymbol{\beta}_0$ indexing the optimal IDR. It is known that a misspecified generative

regression model can result in severe bias in estimating the optimal treatment (Qian and Murphy, 2011; Zhang et al., 2012; Zhao et al., 2012, 2015a). We introduce a robust estimator that accounts for censoring while at the same time circumvents the difficulty of specifying a reliable generative regression model.

Given an IDR $d_{\boldsymbol{\beta}}(\boldsymbol{X})$, the treatment it would recommend to subject i may or may not coincide with the treatment the subject actually received. Even if $A_i = d_{\boldsymbol{\beta}}(\boldsymbol{X}_i)$, we may not observe $T_i^*\{d_{\boldsymbol{\beta}}(\boldsymbol{X}_i)\}$ if the subject is censored. To obtain a consistent estimator for $Q_\tau\{T_i^*\{d_{\boldsymbol{\beta}}(\boldsymbol{X}_i)\}\}$, we adapt the induced missing data framework in Zhang et al. (2012) to time-to-event data. Specifically, we consider an artificial missing data structure with the missing data indicator $R_i(\boldsymbol{\beta}) = \big[A_i d_{\boldsymbol{\beta}}(\boldsymbol{X}_i) + (1-A_i)\{1-d_{\boldsymbol{\beta}}(\boldsymbol{X}_i)\}\big]\Delta_i$. The observed outcome Y_i is equal to the potential outcome $T_i^*\{d_{\boldsymbol{\beta}}(\boldsymbol{X}_i)\}$ only if $R_i(\boldsymbol{\beta}) = 1$. In this framework, the "full data" that we may not completely observe consist of $\{\boldsymbol{X}_i, T_i^*\{d_{\boldsymbol{\beta}}(\boldsymbol{X}_i)\}\}_{i=1}^n$, and the observed data consist of $\big\{\boldsymbol{X}_i, R_i(\boldsymbol{\beta}), R_i(\boldsymbol{\beta})T_i^*\{d_{\boldsymbol{\beta}}(\boldsymbol{X}_i)\}\big\} = \big\{\boldsymbol{X}_i, R_i(\boldsymbol{\beta}), R_i(\boldsymbol{\beta})Y_i\big\}_{i=1}^n$.

Let $\pi_A(\boldsymbol{X}_i) = P(A_i = 1|\boldsymbol{X}_i)$ be the propensity score; and let $G_C\big(t|\boldsymbol{X}, A\big) = P\big(C > t|\boldsymbol{X}, A\big)$ denote the conditional survival function of C given $\{\boldsymbol{X}, A\}$. Let $\pi_i\big(\boldsymbol{\beta}\big) = P\big\{R_i\big(\boldsymbol{\beta}\big) = 1 \mid \boldsymbol{X}_i, T_i^*\big(1\big), T_i^*\big(0\big)\big\}$ be the probability of missingness conditional on the full data. We observe

$$
\begin{aligned}
&\pi_i\big(\boldsymbol{\beta}\big) \\
&= \pi_A(\boldsymbol{X}_i) d_{\boldsymbol{\beta}}\big(\boldsymbol{X}_i\big) P\big[\Delta_i = 1 \mid \boldsymbol{X}_i, T_i^*\big(1\big), T_i^*\big(0\big), A_i = 1\big] \\
&\quad + (1 - \pi_A(\boldsymbol{X}_i))(1 - d_{\boldsymbol{\beta}}\big(\boldsymbol{X}_i\big)) P\big[\Delta_i = 1 \mid \boldsymbol{X}_i, T_i^*\big(1\big), T_i^*\big(0\big), A_i = 0\big] \\
&= \big\{\pi_A(\boldsymbol{X}_i) d_{\boldsymbol{\beta}}\big(\boldsymbol{X}_i\big) + (1 - \pi_A(\boldsymbol{X}_i))(1 - d_{\boldsymbol{\beta}}\big(\boldsymbol{X}_i\big))\big\} G_C\Big(T_i^*\{d_{\boldsymbol{\beta}}\big(\boldsymbol{X}_i\big)\}\big|\boldsymbol{X}, A_i = d_{\boldsymbol{\beta}}\big(\boldsymbol{X}_i\big)\Big).
\end{aligned}
$$

Note that for the complete cases (corresponding to $R_i\big(\boldsymbol{\beta}\big) = 1$), we have $Y_i = T_i^*\{d_{\boldsymbol{\beta}}\big(\boldsymbol{X}_i\big)\}$ and the corresponding

$$
\pi_i\big(\boldsymbol{\beta}\big) = \big[\pi_A(\boldsymbol{X}_i) d_{\boldsymbol{\beta}}\big(\boldsymbol{X}_i\big) + \big(1 - \pi_A(\boldsymbol{X}_i)\big)\big(1 - d_{\boldsymbol{\beta}}\big(\boldsymbol{X}_i\big)\big)\big] G_C\big(Y_i\big|\boldsymbol{X}, A_i\big).
$$

To estimate $\boldsymbol{\beta}_0$, we propose the following two-step estimator. First, we estimate $Q_\tau\big\{T^*\big(d_{\boldsymbol{\beta}}(\boldsymbol{X})\big)\big\}$

by the following inverse probability weighted estimator

$$\widehat{Q}_\tau\big\{T^*\big(d_{\boldsymbol{\beta}}(\boldsymbol{X})\big)\big\} = \arg\min_b \sum_{i=1}^n \frac{R_i\big(\boldsymbol{\beta}\big)}{\widehat{\pi}_i}\rho_\tau\big(Y_i - b\big), \tag{2}$$

where $\widehat{\pi}_i$ is an estimate of $\pi_i(\boldsymbol{\beta})$ and $\rho_\tau\big(u\big) = u\{\tau - I\big(u < 0\big)\}$ is the quantile loss function (Koenker, 2005). By convention, we define $0/0 = 0$. Next, employing the policy-search idea, we estimate $\boldsymbol{\beta}_0$ by

$$\widehat{\boldsymbol{\beta}}_n = \arg\max_{\boldsymbol{\beta}\in\mathbb{B}^o} \widehat{Q}_\tau\big\{T^*\big(d_{\boldsymbol{\beta}}(\boldsymbol{X})\big)\big\}. \tag{3}$$

The above estimator can be computed using the genetic algorithm in R package `rgenoud` (Mebane Jr and Sekhon, 2011). The estimate of the optimal IDR is $d_{\widehat{\boldsymbol{\beta}}_n}\big(\boldsymbol{X}\big) = I\big(\widehat{\boldsymbol{\beta}}_n^T\boldsymbol{X} > 0\big)$.

Remark 1. A key quantity in estimating $\pi_i\big(\boldsymbol{\beta}\big)$ is the conditional survival function of the censoring variable $G_C\big(t|\boldsymbol{X}, A\big) = P\big(C > t|\boldsymbol{X}, A\big)$. There are several approach for estimating $G_C\big(t|\boldsymbol{X}, A\big)$. For clarity of presentation, as in Goldberg and Kosorok (2012) and Jiang et al. (2017a), we assume that $C \perp \{T^*\big(0\big), T^*\big(1\big), A, \boldsymbol{X}\}$ in the theoretical development. This is often satisfied in real applications where administrative censoring occurs. In this case, $G_C\big(t|\boldsymbol{X}, A\big)$ can be estimated by $\widehat{G}_C(\cdot)$, the classical Kaplan-Meier estimator applied to $\{(Y_i, 1-\Delta_i), i = 1, 2, ..., n\}$. When necessary, we can relax the independent censoring assumption to the conditionally independent censoring assumption $C \perp \{T^*\big(0\big), T^*\big(1\big)\}\big|\{\boldsymbol{X}, A\}$ and employs the local Kaplan-Meier estimator. Without loss of generality, we assume that the first n_1 subjects receive treatment $A = 0$, and the other $(n - n_1)$ subjects receive treatment $A = 1$. The local Kaplan-Meier estimator Gonzalez-Manteiga and Cadarso-Suarez (1994) for $G_C\big(\cdot\,|\boldsymbol{X}, A = 0\big)$ is given by

$$\widehat{G}_C\big(\cdot \mid \boldsymbol{X}, A = 0\big) = \prod_{j=1}^{n_1}\Big\{1 - \frac{B_{n_1 j}\big(\boldsymbol{X}\big)}{\sum_{k=1}^{n_1} I\big(C_k \geq C_j\big)B_{n_1 k}\big(\boldsymbol{X}\big)}\Big\}^{\eta_j(t)}, \tag{4}$$

where $\eta_j(t) = I\big(C_j \leq t, \Delta_j = 0\big)$, and $\big\{B_{n_1 k}\big(\boldsymbol{X}\big), k = 1, ..., n_1\big\}$ is a sequence of non-negative weights adding up to 1. A popular choice is the Nadaraya-Watson's type weights for univariate covariate: $B_{nk}\big(X\big) = \big[\sum_{i=1}^n K\big(h_n^{-1}(X - X_i)\big)\big]^{-1}K\big(h_n^{-1}(X - X_i)\big)$, where $K(\cdot)$ is a

positive kernel function and h_n is a sequence of bandwidths converging to zero as $n \to \infty$. We can obtain $\widehat{G}_C\big(\cdot \mid \boldsymbol{X}, A = 1\big)$ similarly. A third approach is to estimate the conditional survival function using a working model, such as the Cox proportional hazards regression model, as investigated in Zhao et al. (2015a).

Remark 2. The Kaplan-Meier estimator $\widehat{G}_C$ in (2) is sometimes unstable at the tail of the distribution. Practically, a simple approach to improve the stability is by employing an artificial censoring technique in Zhou (2006), based on the intuition that any alteration of a random variable's distribution beyond the quantile of interest would have no impact on the quantile. Specifically, assume there exists a large positive constant $M \in \mathbb{R}$ such that $\sup_{\boldsymbol{\beta}} Q_\tau\big\{T^*\big(d_{\boldsymbol{\beta}}\big)\big\} < M$ and $\sup\{t : G_C(t) > 0\} > M$. The first requirement means the largest achievable τth quantile using IDRs belonging to $\mathcal{D}$ is smaller than M; and the second one ensures every data point has a positive probability of not being censored. Note that these conditions are weak, especially if we are interested in lower quantiles. Let $Y^M = Y \wedge M$ and $\Delta^M = \Delta + \big(1 - \Delta\big) I\big(Y \geq M\big)$. Then it is straightforward to show that $\widehat{\boldsymbol{\beta}}_n^M$ obtained using the transformed data set $\big(\boldsymbol{X}_i, A_i, Y_i^M, \Delta_i^M\big)$, $i = 1, \ldots, n$, is the same as $\widehat{\boldsymbol{\beta}}_n$ in (3).

Remark 3. As the optimization problem is nonconcave and nonsmooth, multiple local optimal may exist. Popular algorithms based on derivatives do not work for this challenging setting. At the same time, it is impractical to exhaustively enumerate all possible solutions and pick the best one. In our numerical experiments, we utilize the genetic algorithm in the R package `rgenoud`, which is useful in such a challenging setting when the objective function is nonconcave and the derivatives do not exist. The genetic algorithm (a type of evolutionary algorithm) is inspired from the biological evolution process. In a genetic algorithm, the problem is encoded in a series of bit strings that are manipulated by the algorithm. It is a stochastic, population-based algorithm that searches randomly by mutation and crossover among population members. It is based on searching for the best solutions using inheritance and strengthening of useful features of multiple objects of a specific application in the process of imitation of their evolution. We refer to Mitchell (1998) and Mebane Jr and Sekhon (2011) for more detailed description of the algorithm and other references. In our numerical

experience, the algorithm provides high-quality solutions with desirable statistical properties.

3 Asymptotic theory

In this section, we present two results regarding the asymptotic properties of the estimated optimal IDR with survival data.

- First, we show that the estimated parameter indexing the optimal IDR has nonstandard asymptotics, which is characterized by the cube-root convergence rate and the non-normal limiting distribution.
- Second, we show that under rather weak conditions, which do not require the optimal IDR to be unique at the population or sample level, the theoretically optimal value can be estimated at a near $n^{-1/2}$-rate.

Both results are novel for optimal IDR estimation with time-to-event outcomes. The first result corresponds to a nonstandard estimation problem with both finite-dimensional and infinite-dimensional estimation problem. The second result deals with the challenging setting of exceptional law where the optimal IDR is nonunique.

3.1 Asymptotic distribution of the estimated parameter indexing the optimal IDR

Write $\boldsymbol{X} = (X_1, ..., X_p)^T \equiv (X_1, \tilde{X}^T)^T$. Let $G_C(\cdot)$ denote the survival function of C. Let $F_{T^*(0)}(t|\boldsymbol{X})$ and $F_{T^*(1)}(t|\boldsymbol{X})$ be the cumulative distribution functions of the potential survival times $T^*\big(0\big)$ and $T^*\big(1\big)$, respectively; and let $f_0(t|\boldsymbol{X})$ and $f_1(t|\boldsymbol{X})$ be the corresponding conditional density functions. Given any $\boldsymbol{\beta} \in \mathbb{B}^p$, let $f_{T^*\left(d_{\boldsymbol{\beta}}\right)}(\cdot)$ denote the marginal density function of the distribution of the potential survival time $T^*(d_{\boldsymbol{\beta}}\big(\boldsymbol{X}\big))$.

To avoid complications irrelevant to the main results of the paper, we consider data collected from a randomized study where $\pi_A(\boldsymbol{X}_i) = 0.5$, but the results can be extended to

observational data under mild assumptions. The estimator $\widehat{Q}_\tau\big(T^*\big(d_{\boldsymbol{\beta}}\big)\big)$ in (2) simplifies to

$$\widehat{Q}_\tau\big(\boldsymbol{\beta};\widehat{G}_C\big) = \arg\min_b \sum_{i=1}^{n} \frac{R_i\big(\boldsymbol{\beta}\big)\ \rho_\tau\big(Y_i - b\big)}{0.5\widehat{G}_C\big(Y_i\big)}, \tag{5}$$

where $\widehat{G}_C(\cdot)$ is the classical Kaplan-Meier estimator of $G_C(\cdot)$. We then estimate $\boldsymbol{\beta}_0$ by $\widehat{\boldsymbol{\beta}}_n = \arg\max_{\boldsymbol{\beta}\in\mathbb{B}^o}\widehat{Q}_\tau\big(\boldsymbol{\beta};\widehat{G}_C\big)$. Write $\widehat{\boldsymbol{\beta}}_n = (\widehat{\beta}_{n1}, \widetilde{\boldsymbol{\beta}}_n^T)^T$, where $\widehat{\beta}_{n1}$ satisfies the identifiability condition $|\widehat{\beta}_{n1}| = 1$. In the proof of Theorem 1 in the online supplement, it was shown that $\widehat{\boldsymbol{\beta}}_n$ is consistent for $\boldsymbol{\beta}_0$. Hence, we have $\widehat{\beta}_{n1} = \beta_{01}$ with probability approaching one. Theorem 1 below states the nonstandard convergence rate and non-normal limiting distribution of $\widehat{\boldsymbol{\beta}}_n$.

Theorem 1. *Suppose conditions (C1)-(C4) are satisfied. Then as $n \to \infty$,*

$$n^{1/3}\big(\widetilde{\boldsymbol{\beta}}_n - \widetilde{\boldsymbol{\beta}}_0\big) \to \arg\max_t \big\{\Psi(t) + \mathbb{W}(t)\big\} \tag{6}$$

in distribution, where $\Psi(t)$ is a deterministic function whose form is given in (17) of the online supplement and $\mathbb{W}(t)$ is a mean-zero Gaussian process with covariance function given in (19) in the online supplement.

The proof of Theorem 1 is given in the online supplement. Theoretical analysis of the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_n$ in (5) is challenging, as it is defined via a bilevel optimization problem. The proof involves reformulating $\widehat{\boldsymbol{\beta}}_n$ as an M-estimator with a nonsmooth objective function that has two nuisance parameters: a finite dimensional nuisance parameter m_0 and an infinite dimensional nuisance parameter $G_C(\cdot)$. The nonstandard asymptotics arise from the so-called *sharp-edge effect*, see Kim and Pollard (1990) for an informative example of the shorth estimator that illustrates this phenomenon. It is worth noting that the theory in Kim and Pollard (1990) can only handle a finite dimensional nuisance parameter, hence is not applicable in our setting.

Remark 4. Our results are related to recent work on non-standard estimation problem in Banerjee and McKeague (2007), Sen et al. (2010), Matsouaka et al. (2014), Wang et al. (2018), Shi et al. (2018), Patra et al. (2018) and Banerjee et al. (2019). However, none of the above work involves an infinite-dimensional nuisance parameter as we face in the current set-

ting. In fact, our estimation method involves both finite-dimensional and infinite-dimensional nuisance parameters, the role of the latter is for estimating the censoring distribution. Due to the nature of nonstandard asymptotics, the problem is different from and much harder than regular M-estimation problem with infinite-dimensional nuisance parameters. Advanced empirical process techniques from van der Vaart and Wellner (1996), Kosorok (2008) and Delsol and Van Keilegom (2020) were adapted here to help deal with the theoretical challenges.

3.2 Estimating the optimal value with possibly non-unique optimal IDR

Besides the optimal IDR itself, a quantity of interest is the optimal value, defined as

$$V_{opt} = \sup_{\boldsymbol{\beta}\in\mathbb{B}^o} Q_\tau\big\{T^*\big(d_{\boldsymbol{\beta}}(\boldsymbol{X})\big)\big\}. \tag{7}$$

This quantity is the maximally achievable marginal τ-th quantile of the potential distribution of all IDRs in the given class of candidate rules. It is an important measure of the performance of the optimal IDR. A natural estimator of this quantity is $\widehat{V}_n = \widehat{Q}_\tau\big(\widehat{\boldsymbol{\beta}}_n; \widehat{G}_C\big)$.

Theorem 2 shows that under rather weak conditions, $\widehat{V}_n$ provides a near $n^{-1/2}$-rate estimator for V_{opt}. To avoid the requirement of a unique optimal IDR, the derivation of this result is based on recognizing that the following equivalent expressions:

$$V_{opt} = \sup\Big\{v : \sup_{\boldsymbol{\beta}\in\mathbb{B}^o} Pg\big(\cdot, \boldsymbol{\beta}, v, G_C\big) \geq 1-\tau\Big\}, \tag{8}$$

$$\widehat{V}_n = \sup\Big\{v : \sup_{\boldsymbol{\beta}\in\mathbb{B}^o} P_n g\big(\cdot, \boldsymbol{\beta}, v, \widehat{G}_C\big) \geq 1-\tau\Big\}, \tag{9}$$

where

$$g\big(\cdot, \boldsymbol{\beta}, v, G_C\big) = \frac{R(\boldsymbol{\beta})}{0.5G_C(Y)} I(Y - v > 0). \tag{10}$$

The function $g\big(\cdot, \boldsymbol{\beta}, v, G_C\big)$ is motivated by the first-order optimization condition of the estimator $\widehat{Q}_\tau\big(\boldsymbol{\beta}; \widehat{G}_C\big)$. A careful inspection reveals that (7) and (8) are equivalent. To see this, we observe that for a given $\boldsymbol{\beta}$, $g\big(\cdot, \boldsymbol{\beta}, v, G\big)$ is a monotonically decreasing function of v,

and $Pg\big(\cdot, \boldsymbol{\beta}, v, G\big) \geq 1-\tau$ when $v \leq Q_\tau\big\{T^*\big(d_{\boldsymbol{\beta}}\big)\big\}$. Correspondingly, $\boldsymbol{\beta}_0$ is the parameter indexing the IDR that achieves this V_{opt}.

Theorem 2. *Suppose conditions (C1) is satisfied. We have* $\widehat{V}_n = V_{opt} + o_p(n^{-1/2+\gamma_0})$, *for an arbitrary* $\gamma_0 > 0$.

Remark 5. It is worth emphasizing that Theorem 2 requires much weaker conditions than Theorem 1 does. In particular, it does not require the optimal IDR to be unique at the population or sample level. This corresponds to a well known challenging situation where there exists a subpopulation who responds similarly to the two treatment options. If one is willing to assume unique optimal IDR, then the above result can be strengthened to parametric convergence rate, i.e., $n^{-1/2}$ rate.

3.3 Smoothed resampling inference

Statistical inference for $\widetilde{\boldsymbol{\beta}}_0$ is challenging due to the nonstandard asymptotic distribution. A natural idea is to use bootstrap. However, the standard nonparametric bootstrap procedure is generally inconsistent for cube-root M-estimators (e.g., Abrevaya and Huang (2005), Léger and MacGibbon (2006)) even for the relatively simpler setting without nuisance functions. As a remedy, m-out-of-n bootstrap (Bickel et al., 2012), which draws subsamples of size m from the original sample of size n with replacement, has been shown to be consistent for M-estimators with a cube root convergence rate in some settings (Delgado et al., 2001). Theoretically, m depends on n, tends to infinity with n, and satisfies $m = o(n)$. Practically, choosing an optimal m is not a simple task. Several data-driven approaches for selecting m were investigated but require intensive computation (e.g.,Banerjee and McKeague (2007), Bickel and Sakov (2008), Chakraborty et al. (2013), Qian et al. (2021)).

In this subsection, we consider an alternative smoothed resampling-based procedure which is computationally more convenient. This approach is motivated by the alternative expression of $\boldsymbol{\beta}_0$ (see the derivation of Lemma 1 in the online supplement), given by

$$\boldsymbol{\beta}_0 = \arg\max_{\boldsymbol{\beta} \in \mathbb{B}^o} Pg(\cdot, \boldsymbol{\beta}, V_{opt}, G_C),$$

where V_{opt} is the optimal value and $g\big(\cdot, \boldsymbol{\beta}, v, G_C\big)$ is defined in (10). That is, $\boldsymbol{\beta}_0$ is the parameter indexing the IDR that achieves the optimal value V_{opt}. This naturally leads to an alternative representation of $\widehat{\boldsymbol{\beta}}_n$, given by $\widehat{\boldsymbol{\beta}}_n = \arg\max_{\boldsymbol{\beta}\in\mathbb{B}^o} n^{-1}\sum_{i=1}^n g\left(\cdot, \boldsymbol{\beta}, \widehat{V}_n, \widehat{G}_C\right)$.

To implement the smoothed resampling-based inference, we first obtain the estimator $\widehat{\boldsymbol{\beta}}_n$ and then estimate the optimal value function by $\widehat{V}_n = \arg\min_b \sum_{i=1}^n \frac{R_i\left(\widehat{\boldsymbol{\beta}}_n\right)\ \rho_\tau\left(Y_i - b\right)}{0.5\widehat{G}_C\left(Y_i\right)}$. Motivated by Wu and Wang (2021) for mean-optimal treatment regime with complete data, we consider the following smoothed estimator

$$\overline{\boldsymbol{\beta}}_n = \arg\max_{\boldsymbol{\beta}\in\mathbb{B}^o} \frac{1}{n}\sum_{i=1}^n (2A_i - 1)\frac{\Delta_i I\left(Y_i > \widehat{V}_n\right)}{G_C\left(Y_i\right)} K\left(\frac{\boldsymbol{X}_i^T\boldsymbol{\beta}}{h_n}\right),$$

where $K(\cdot)$ is a kernel function and h_n is a bandwidth. The kernel function $K(\cdot)$ is only required to satisfy some general conditions, for example, we can take it to be the cumulative distribution function of the standard normal distribution. Replacing the indicator function in the treatment regime by the kernel function helps alleviates the sharp edge effect. Write $\overline{\boldsymbol{\beta}}_n = (\overline{\beta}_{n1}, \overline{\widetilde{\boldsymbol{\beta}}}_n^T)^T$ Similarly as in Wu and Wang (2021), it is expected that $\sqrt{nh}\left(\overline{\widetilde{\boldsymbol{\beta}}} - \widetilde{\beta}_0\right)$ is asymptotically normal. Note that $\overline{\boldsymbol{\beta}}_n$ minimizes the loss function $-n^{-1}\sum_{i=1}^n (2A_i - 1)\frac{\Delta_i I\left(Y_i > \widehat{V}_n\right)}{G_C(Y_i)} K\left(\frac{\boldsymbol{X}_i^T\boldsymbol{\beta}}{h_n}\right)$, which is a smoothed estimator of a weighted misclassification error. We choose h_n by five-fold cross-validation based on this loss function.

The asymptotic covariance matrix is complex and involves unknown counter-factual distributions. For inference, we consider the following perturbed smoothed estimator

$$\overline{\boldsymbol{\beta}}_n^* = \arg\max_{\boldsymbol{\beta}\in\mathbb{B}^o} \frac{1}{n}\sum_{i=1}^n \xi_i (2A_i - 1)\frac{\Delta_i I\left(Y_i > \widehat{V}_n\right)}{G_C\left(Y_i\right)} K\left(\frac{\boldsymbol{X}_i^T\boldsymbol{\beta}}{h_n}\right),$$

where $\xi_1, \cdots, \xi_n$ are positive random weights independent of the data, with mean one and variance one. To obtain the bootstrap distribution of $\overline{\boldsymbol{\beta}}_n^*$, we repeatedly generate independent random weights and solve for the smoothed estimator. Write $\overline{\boldsymbol{\beta}}_n = (\overline{\beta}_{n1}, \overline{\beta}_{n2}, \ldots, \overline{\beta}_{np})$, and $\overline{\boldsymbol{\beta}}_n^* = (\overline{\beta}_{n1}^*, \overline{\beta}_{n2}^*, \ldots, \overline{\beta}_{np}^*)$, where $|\overline{\beta}_{n1}^*| = 1$. For $j = 2, \ldots, p$, let $\eta_j^{*(\alpha/2)}$ and $\eta_j^{*(1-\alpha/2)}$ be the $(\alpha/2)$-th and $(1-\alpha/2)$-th quantile of the bootstrap distribution of $(nh_n)^{1/2}(\overline{\beta}_{nj}^* - \overline{\beta}_{nj})$, respectively, where α is a small positive number. We can estimate $\eta_j^{*(\alpha/2)}$ and $\eta_j^{*(1-\alpha/2)}$ from

a large number of bootstrap samples. An asymptotic $100(1-\alpha)\%$ bootstrap confidence interval for β_{0j}, $j = 2, \ldots, p$, is given by $\left\{\overline{\beta}_{nj} - (nh_n)^{-1/2}\eta_j^{*(1-\alpha/2)}, \overline{\beta}_{nj} - (nh_n)^{-1/2}\eta_j^{*(\alpha/2)}\right\}$..

4 Estimation of optimal dynamic IDR with censored data.

In this section, we consider the extension to the dynamic decision problem which involves multiple decision points. The decision at a later stage can depend on baseline covariates, treatment history, and intermediate variables such as how the subject responds to earlier treatment(s). For survival data, complications arise as the subject may be censored anytime before the end of the study, which results in an incomplete trajectory of treatments.

4.1 Potential outcome framework

For clarity, we focus on a two-stage dynamic decision problem with random right censoring. We consider a setup similar as Jiang et al. (2017a) but will define the potential outcome more carefully. At the beginning of a study (time point 0), baseline covariates $\boldsymbol{X}_1$ of patients would be collected, and each of them is assigned one of two first-stage treatment options, say A_1 and A_2. Then the second stage starts from a prespecified time s with $s > 0$. Additional intermediate covariates $\boldsymbol{X}_2$ reflecting the reaction to first stage treatment up to time s would be collected if applicable, and those subjects who remain at risk at time s is assigned one of two first-stage treatment options, say B_1 and B_2. $\{B_1, B_2\}$ may not overlap with $\{A_1, A_2\}$. For example, in making decisions for cancer patients, $\{A_1, A_2\}$ could be induction treatments, while $\{B_1, B_2\}$ may represent maintenance treatment or salvage treatment.

Similarly as in the one-stage setting, the potential survival time is defined when censoring is absent, and we would like to estimate the optimal dynamic IDR with a criterion based on this potential survival time when the real data is complicated by censoring. Let D_i denote the random treatment at stage i when the subject is eligible. Note that D_2 may not exist if the patient is not at risk at time s. Consider a sequential IDR $\boldsymbol{d} = (d_1, d_2)$, where $d_1(\boldsymbol{X}_1) \in \{A_1, A_2\}$ and $d_2(\boldsymbol{X}_1, D_1, \boldsymbol{X}_2) \in \{B_1, B_2\}$. A subject is considered to be consistent

with $\boldsymbol{d}$ if he/she receives a first treatment D_1 that equals $d_1(\boldsymbol{X}_1)$ and a second treatment D_2 which equals $d_2(\boldsymbol{X}_1, D_1, \boldsymbol{X}_2)$ (full compliance); *or* receives treatment D_1 complying with the rule d_1 at stage one but does not survive long enough to be eligible for stage two treatment. Let $\dot{R}(d_1) = I\Big(\dot{T}\,(d_1, \emptyset) > s\Big)$ indicate the subject's eligibility status for stage two treatment when complying with rule d_1, where $\dot{T}\,(d_1, \emptyset)$ is shorthand notation of $\dot{T}(d_1(\boldsymbol{X}_1), \emptyset)$, which represents the potential survival time if the subject receives d_1 without stage-two action. Let $\dot{T}\,(d_1, d_2)$ be the potential survival time if the subject receives the full sequence of treatments (d_1, d_2). Implicitly, $\dot{T}\,(d_1, d_2) > s$. Let $T^*(\boldsymbol{d})$ be the potential survival time if the subject is consistent with the treatment sequence $\boldsymbol{d}$. We can write

$$T^*(\boldsymbol{d}) = \dot{T}\,(d_1, \emptyset)\big(1 - \dot{R}(d_1)\big) + \dot{T}\,(d_1, d_2)\dot{R}(d_1). \tag{11}$$

We are interested in estimating the optimal sequential decision $\boldsymbol{d} = (d_1, d_2)$ in some class $\mathcal{D}$, that is, $\boldsymbol{d}_{opt} = \arg\max_{\boldsymbol{d}\in\mathcal{D}} Q_\tau(\boldsymbol{d})$.

Define $H_1 = \{\boldsymbol{X}_1\}$, and define $H_2^*(d_1) = \{\boldsymbol{X}_1, d_1, \boldsymbol{X}_2^*(d_1)\}$, where $\boldsymbol{X}_2^*(d_1)$ denotes the potential intermediate information between decision 1 and decision 2 had the subject started with treatment $d_1(\boldsymbol{X}_1)$ and given $\dot{R}(d_1) = 1$. Denote the set of potential outcomes as

$$O^*(\boldsymbol{d}) = \Big\{\dot{T}\,(d_1, \emptyset),\ \dot{R}(d_1),\ \dot{R}(d_1)\boldsymbol{X}_2^*(d_1),\ \dot{R}(d_1)d_2\big(H_2^*(d_1)\big),\ T^*(\boldsymbol{d})\Big\}.$$

4.2 Robust estimation of optimal IDR

When censoring is absent, for a given subjects, we would observe the first stage treatment D_1 and the corresponding $\dot{R}(D_1)$. The consistency assumption for causal inference, similar as in the one-stage setting, ensures that the observed survival time T would satisfy

$$T = \begin{cases} \sum_{j\in\{1,2\}} I\{D_1 = A_j\}\dot{T}(A_j, \emptyset), & \text{if } \dot{R} = 0 \\ \sum_{j\in\{1,2\},\ k\in\{1,2\}} I\{D_1 = A_j, D_2 = B_k\}\dot{T}\big(A_j, B_k\big) & \text{if } \dot{R} = 1. \end{cases} \tag{12}$$

Due to censoring, we may not observe T. Denote the actually observed survival time under possible censoring as $Y = \min\{T, C\}$, and let $\Delta = I(T \le C)$ be the censoring

indicator. Further, let $\Gamma = I(C > s)$ denote whether censoring occurs in the first stage. As a result of censoring, only those subjects that survived longer than s and are not censored before s are eligible for the second-stage treatment, for whom the trajectory observed up to time s is $H_2 = \{\boldsymbol{X}_1, D_1, \boldsymbol{X}_2\}$. When the trial ends, the observed data is

$$\left\{\boldsymbol{X}_{i1}, D_1,\ \dot{R}_i,\ \Gamma_i,\ \dot{R}_i\Gamma_i\boldsymbol{X}_{i2},\ \dot{R}_i\Gamma_i D_2,\ Y_i,\ \Delta_i\right\},\ \text{for } i = 1, \ldots, n. \tag{13}$$

Based on the observed data, our goal is to estimate the optimal sequential decision rule within a given class $\mathcal{D}$. Extending the one-stage formulation, we consider sequential IDR of the form $\boldsymbol{d}_{\boldsymbol{\xi}} = \{d_{1,\boldsymbol{\beta}}(H_1), d_{2,\boldsymbol{\zeta}}(H_2)\}$, where $d_{1,\boldsymbol{\beta}}(H_1) = d_{1,\boldsymbol{\beta}}(\boldsymbol{X}_1) = I(\boldsymbol{X}_1^T\boldsymbol{\beta} > 0)$, $d_{2,\boldsymbol{\zeta}}(H_2) = d_{2,\boldsymbol{\zeta}}(\boldsymbol{X}_1, D_1, \boldsymbol{X}_2) = I(H_2^T\boldsymbol{\zeta} > 0)$ and $\boldsymbol{\xi} = (\boldsymbol{\beta}^T, \boldsymbol{\zeta}^T)^T$. Without loss of generality, we assume that if $d_{1,\boldsymbol{\beta}}(\boldsymbol{X}_1) = 0$, then the recommended first-stage treatment is A_1, otherwise is A_2; and if $d_{2,\boldsymbol{\zeta}}(H_2) = 0$, then the recommended second-stage treatment is B_1, otherwise is B_2. For identifiability, we assume $\boldsymbol{\beta}$ and $\boldsymbol{\zeta}$ satisfy $|\beta_1| = 1$ and $|\zeta_1| = 1$. Hence $\mathcal{D} = \{\boldsymbol{d}_{\boldsymbol{\xi}} : \boldsymbol{\xi} \in \widetilde{\mathbb{C}}\}$ where $\widetilde{\mathbb{C}} = \{-1, 1\} \times \widetilde{\mathbb{B}} \times \{-1, 1\} \times \widetilde{\mathbb{Z}}$, with $\widetilde{\mathbb{B}}$ being a compact subset of $\mathbb{R}^{p_1-1}$, $\widetilde{\mathbb{Z}}$ being a compact subset of $\mathbb{R}^{p_2-1}$; p_1 is the dimension of $\boldsymbol{X}_1$, and p_2 is the dimension of $(\boldsymbol{X}_1^T, D_1, \boldsymbol{X}_2^T)^T$. The parameter $\boldsymbol{\xi}_0$ indexing the optimal dynamic IDR in $\mathcal{D}$ is defined by:

$$\boldsymbol{\xi}_0 = \arg\max_{\boldsymbol{\xi} \in \widetilde{\mathbb{C}}} Q_\tau\{T^*(\boldsymbol{d}_{\boldsymbol{\xi}})\},$$

where $Q_\tau\{\cdot\}$ denotes the marginal τth quantile ($0 < \tau < 1$), and $T^*(\boldsymbol{d}_{\boldsymbol{\xi}})$ is obtained by setting $\boldsymbol{d} = \boldsymbol{d}_{\boldsymbol{\xi}}$ in (11). To extend the policy-search method to estimate $\boldsymbol{\xi}_0$, we define

$$\widetilde{R}(\boldsymbol{d}_{\boldsymbol{\xi}}) = \Delta I(D_1 = d_{1,\boldsymbol{\beta}}(\boldsymbol{X}_1))\ [I(Y \le s) + I(Y > s)I(D_2 = d_{2,\boldsymbol{\zeta}}(H_2))].$$

For subjects with $\widetilde{R}(\boldsymbol{d}_{\boldsymbol{\xi}}) = 1$, we observe the potential survival time of interest, that is, $Y = T^*(\boldsymbol{d}_{\boldsymbol{\xi}})$.

For simplicity in presentation, we assumed that the data are from a sequential multiple assignment randomized trial (SMART, (Lavori and Dawson, 2000; Murphy, 2008)), where the randomization probabilities at each stage are known by design. That is, at stage one, $P(D_1 = A_2) = 1 - P(D_1 = A_1) = \pi_1$, while at stage two $P(D_2 = B_2 | Y_i > s, C_i >$

$s) = 1 - P(D_2 = B_1 | Y_i > s, C_i > s) = \pi_2$. Given $\boldsymbol{d} = (d_1, d_2) \in \mathcal{D}$, let $\pi_{d_1}(\boldsymbol{X}_{i1}) = \pi_1 d_1(\boldsymbol{X}_{i1}) + (1-\pi_1)\{1 - d_1(\boldsymbol{X}_{i1})\}$ denote the probability of compliance to d_1 at stage one; and let $\pi_{d_2}(H_{i2}) = \pi_2 d_2(H_{i2}) + (1-\pi_2)\{1 - d_2(H_{i2})\}$ denote the probability of compliance to d_2 at stage two, given that stage one's target potential data is observed and also $Y > s, C > s$. Overall, the probability to observe $T^*(\boldsymbol{d})$ is

$$\widetilde{w}_{\boldsymbol{d},i} = P\left(\widetilde{R}_i(\boldsymbol{d}) = 1 \mid \boldsymbol{X}_{i1}, O_i^*\big(\boldsymbol{d}\big)\right) = \pi_{d_1}(\boldsymbol{X}_{i1}) G_C(Y_i)\big\{I(Y_i \le s) + \pi_{d_2}(H_{i2}) I(Y_i > s)\big\}. \tag{14}$$

We estimate $Q_\tau\big\{T^*(\boldsymbol{d})\big\}$ by

$$\widehat{Q}_\tau\big(T^*(\boldsymbol{d}); \widehat{G}_C\big) = \arg\min_b \sum_{i=1}^n \frac{\widetilde{R}_i(\boldsymbol{d}) \rho_\tau\big(Y_i - b\big)}{\widehat{w}_{\boldsymbol{d},i}},$$

where $\widehat{w}^{(2)}_{\boldsymbol{d},i}$ is obtained by plugging in the Kaplan-Meier estimator for G_C in (14). For brevity, we use the shorthand notation $\widehat{Q}_\tau\big(\boldsymbol{\xi}; \widehat{G}_C\big)$ for $\widehat{Q}_\tau\big(T^*(\boldsymbol{d}_{\boldsymbol{\xi}}); \widehat{G}_C\big)$.

Let L denote the end of the study. Assume there exists a constant $\eta > 0$ such that $G_C(L) > \eta > 0$. Furthermore, assume C has a continuously differentiable density function which is bounded away from infinity on $(0, L)$. Also, $m_0' \triangleq \sup_{\boldsymbol{\xi} \in \widetilde{\mathbb{C}}} Q_\tau\{T^*(\boldsymbol{d}_{\boldsymbol{\xi}})\} < L$. Consider an arbitrary treatment sequence $\boldsymbol{d} = (d_1, d_2)$, with $d_1(H_1) \in \{A_1, A_2\}$ and $d_2(H_2) \in \{B_1, B_2\}$. Marginally, $\dot{T}\,(d_1, \emptyset)$ and $\dot{T}\,(d_1, d_2)$ have continuous distributions with continuously differentiable density functions. $\forall \boldsymbol{\xi} \in \widetilde{\mathbb{C}}$, let $f_{T^*(\boldsymbol{d}_{\boldsymbol{\xi}})}(\cdot)$ denote the marginal density function of the distribution of the potential survival time $T^*(d_{\boldsymbol{\xi}})$. There exist positive constants κ_1 and δ, such that $\inf_{\boldsymbol{\xi} \in \widetilde{\mathbb{C}}} \inf_{|m - m_0'| \le \delta} f_{T^*(d_{\boldsymbol{\xi}})}(m) \ge \kappa_1$.

The following lemma states the consistency of $\widehat{Q}_\tau\big(\boldsymbol{\xi}; \widehat{G}_C\big)$ for the marginal quantile of $T^*(\boldsymbol{d}_{\boldsymbol{\xi}})$.

Lemma 1. *For all $d_{\boldsymbol{\xi}} \in \mathcal{D}$, we have $\widehat{Q}_\tau\big(\boldsymbol{\xi}; \widehat{G}_C\big) \to Q_\tau\{T^*(\boldsymbol{d}_{\boldsymbol{\xi}})\}$ in probability.*

Hence, an estimator of the parameter $\boldsymbol{\xi}_0$ is

$$\widehat{\boldsymbol{\xi}}_n = \arg\max_{\boldsymbol{\xi} \in \widetilde{\mathbb{C}}} \widehat{Q}_\tau\big\{T^*\big(\boldsymbol{d}_{\boldsymbol{\xi}}\big); \widehat{G}_C\big\}. \tag{15}$$

The optimal value function, the maximally achievable marginal τ-th quantile of the potential

outcome distribution considering all IDRs in $\mathcal{D}$, is given by $V_{opt} = \sup_{\boldsymbol{\xi}\in\widetilde{\mathbb{C}}} Q_\tau\{T^*(d_{\boldsymbol{\xi}})\}$. An estimator of this quantity is $\widehat{V}_n = \widehat{Q}_\tau(\widehat{\boldsymbol{\xi}}_n, \widehat{G}_C)$. Similarly as the one-stage setting, we can re-express $\widehat{V}_n$ as

$$\widehat{V}_n = \sup\left\{m : \sup_{\boldsymbol{\xi}\in\widetilde{\mathbb{C}}} n^{-1}\sum_{i=1}^{n} g\left(\cdot, \boldsymbol{\xi}, m, \widehat{G}_C\right)\right\} \tag{16}$$

where $g(\cdot, \boldsymbol{\xi}, m, G_C) = \frac{\widetilde{R}(\boldsymbol{d})I(Y-m>0)}{\widetilde{w}_{\boldsymbol{d}}}$. The following theorem shows that in the dynamic setting, the optimal value function can be estimated with a near parametric rate.

Theorem 3. *For the estimator $\widehat{V}_n$ defined in (16), we have $\widehat{V}_n = V_{opt} + o_p(n^{-1/2+\gamma_0})$ for an arbitrary $\gamma_0 > 0$.*

It is worth noting that the above result does not require the optimal IDR to be unique. Similar nonregular asymptotic distribution for $\widehat{\boldsymbol{\xi}}_n$ can also be established using the same idea as in Section 3 but more complex notations.

Remark 6. The method we propose for the dynamic setting is different from the Q-learning approach, which searches for optimal treatment regimes starting from the last stage and moving backward. The Q-learning approach was extended to censored data by Goldberg and Kosorok (2012). There exist several distinct differences between these two methods. First, the proposed method is model-free in the sense that it does not require to specify an outcome regression model. The Q-learning approach is model-based and requires a survival time model that incorporates both the covariate effects and treatment-covariate interaction effects. Second, the proposed method considers a quantile-optimal criterion while Goldberg and Kosorok (2012) adopts a restricted mean criterion. Finally, from a theoretical perspective, this work focuses on the statistical properties of the estimated parameters indexing the optimal IDR while Goldberg and Kosorok (2012) focuses on the finite sample bound of the generalization error of the estimated optimal IDR.

5 Numerical studies

5.1 Monte Carlo simulations

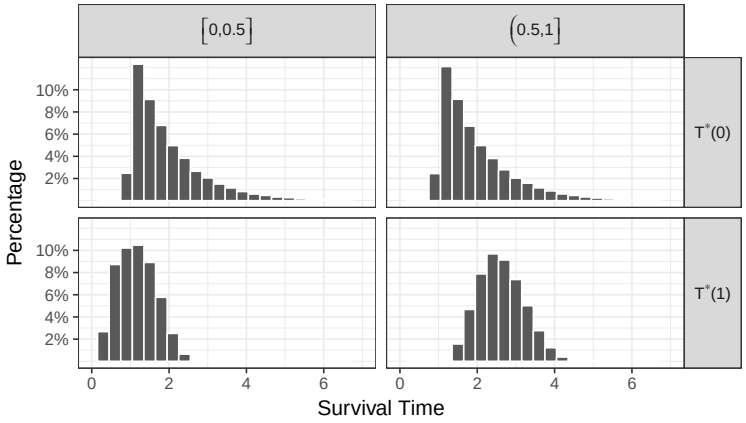
We report simulation results for three different settings. In the first example, we estimate one-stage optimal treatment under random censoring; in the second example, we estimate one-stage optimal treatment under covariate-dependent censoring; while in the third example, we consider a two-stage dynamic optimal IDR estimation problem.

Example 1 (random censoring). We generate the random sample $\{\boldsymbol{X}_i, A_i, Y_i, \Delta_i\}$, $i = 1, 2, ..., n$, from the model: $X_1 \sim U(0,1)$, $T^*\big(0\big)\big|X_1 \sim$ Weibull(shape $= 1$, scale $= 1) + 1$, $T^*\big(1\big)\big|X_1 \sim$ Weibull(shape $= 3$, scale $= 0.5+X_1)+2X_1$, $A\big|X, T^*\big(0\big), T^*\big(1\big) \sim$ Bernoulli(0.5). The response variable in the absence of censoring is generated by $T = T^*\big(0\big)\,(1-A) + T^*\big(1\big)A$. The censoring time C has a constant density function 0.22 on $(0,2)$ and a constant density function 0.07 on $[2,10)$. The observed response is $Y = \min\{T, C\}$ and the censoring indicator is $\Delta = I\,\{T \leq C\}$. This setup achieves an overall censoring rate of 35%.

To illustrate the heterogeneous treatment effects, we split X_1 into two strata: $[0, 0.5]$ and $(0.5, 1]$. Figure 1 displays the histograms of $T^*\big(0\big)$ and $T^*\big(1\big)$ in each stratum. This plot provides strong evidence that X_1 has a qualitative interaction with the treatment. Intuitively, the optimal IDR should depend on X_1. We will apply the proposed method to estimate the quantile-optimal IDR for $\tau = 0.25$ and $\tau = 0.5$, respectively. We consider the following class of IDRs $\mathcal{D} = \left\{d_{\boldsymbol{\beta}}\big(\boldsymbol{X}\big) = \mathrm{I}\left(X_1\beta_1 + \beta_2 > 0\right) : |\beta_1| = 1, \beta_2 \in \mathbb{R}\right\}$. Denote the parameter indexing the τth quantile optimal IDR in $\mathcal{D}$ by $\boldsymbol{\beta}_0^{(\tau)}$. For each τ, we use a large Monte Carlo data set of size $n = 10^7$ to estimate $\boldsymbol{\beta}_0^{(\tau)}$ and the τth quantile of the potential outcome in the above class of IDRs (denoted by Q_τ) and treat the results as population parameter values, see Table 1. Consider, for example, the row corresponding to $\tau = 0.5$. We apply the 0.5-quantile optimal IDR to assign treatment in a large independent Monte Carlo sample. Assume everyone in the population follows the recommended treatment and records his/her outcome. The median of the potential outcome distribution is 2.258, the first quartile of the potential outcome distribution is 1.587.

We compare the proposed estimator (denoted by New) with the naive estimator (denoted

Figure 1: Histograms of $T^*(0)$ and $T^*(1)$ stratified by X_1



τ	$\beta_{01}^{(\tau)}$	$\beta_{02}^{(\tau)}$	$Q_{0.25}$	$Q_{0.5}$
0.25	1	-0.428	1.658	2.215
0.50	1	-0.552	1.587	2.258

Table 1: Example 1: Parameters indexing the quantile-optimal IDRs ($\tau = 0.25$ and 0.5) in $\mathcal{D}$ and the τth quantile of the potential outcome distribution (denoted by Q_τ), based on a Monte Carlo experiment ($n = 10^7$).

by Naive), which ignores censoring and pretends all observations are complete (Wang et al., 2018). We conduct the simulation experiment with 400 replications for sample size $n = 300$, 500, and 1000. In this experiment, we observed that New always correctly estimates the sign of β_{01} for both $\tau = 0.25$ and $\tau = 0.5$, while Naive has 4%, 2%, and 1% error rate for $n = 300, 500$ and 1000 respectively in estimating the sign of β_{01} for $\tau = 0.25$ (0% error rate for $\tau = 0.5$). Consideration this phenomenon, we conservatively compare the estimates for β_{02} for the two methods in cases where $\hat{\beta}_{01} = 1$. Table 2 summarizes the bias (with standard deviation in the parenthesis) of New and Naive for estimating $\boldsymbol{\beta}_0^{(\tau)}$ and Q_τ for different combinations of τ and n. To estimate Q_τ, New plugs $\widehat{\boldsymbol{\beta}}_n$ into the formula of $\widehat{Q}_\tau\left(\cdot; \widehat{G}_C\right)$ in (5); and Naive does plugs-in similarly pretending all observation were complete. We observe that New has satisfactory performance, while Naive exhibits substantial bias for estimating both $\boldsymbol{\beta}_0^{(\tau)}$ and Q_τ.

Finally, we demonstrate the smoothed resampling procedure in Section 3.3 for inference. We consider 90% and 95% confidence intervals for β_{02} for $\tau = 0.25$ and 0.5, respectively. The empirical coverage probabilities and average confidence interval lengths are reported in

τ	n	New		Naive	
		$\beta_{02}^{(\tau)}$	Q_τ	$\beta_{02}^{(\tau)}$	Q_τ
0.25	300	0.005(0.066)	0.056(0.113)	-0.025(0.204)	-0.555(0.057)
	500	-0.001(0.054)	0.027(0.082)	-0.043(0.213)	-0.568(0.050)
	1000	0.001(0.043)	0.020(0.055)	-0.048(0.202)	-0.585(0.031)
0.50	300	0.002(0.098)	0.048(0.124)	0.129(0.082)	-0.618(0.078)
	500	0.003(0.080)	0.023(0.101)	0.122(0.064)	-0.617(0.065)
	1000	-0.002(0.051)	0.022(0.061)	0.134(0.051)	-0.649(0.049)

Table 2: Bias (with standard deviation in the parenthesis) of New and Naive for estimating $\boldsymbol{\beta}_0^{(\tau)}$ and Q_τ for Example 1.

Table 3 for $n = 1000$ based on 400 bootstrap samples. The observed empirical coverage probabilities are close to the nominal levels with reasonable lengths.

		90% CI		95% CI	
n	τ	coverage	length	coverage	length
500	$\tau = 0.5$	0.89	0.17	0.92	0.21
	$\tau = 0.25$	0.91	0.29	0.95	0.34
1000	$\tau = 0.5$	0.88	0.13	0.93	0.16
	$\tau = 0.25$	0.88	0.14	0.93	0.16

Table 3: Confidence intervals for β_{02} using smoothed resampling

Example 2 (covariate-dependent censoring). Let $\boldsymbol{X}_i = (X_{i1}, 1, X_{i2})^T$, where X_{i1}, X_{i2} are independent Uniform $(0, 1)$ random variables. The binary treatment A_i is independent of $\boldsymbol{X}_i$ and satisfies $P(A_i = 1) = 0.5$. The distribution of censoring variable C_i is

$$C_i = \begin{cases} 4 + (2 - X_{i1})\,\omega_i, & \text{if } A_i = 0 \\ 2 + I\left(X_{i1} < 0.5 \cup X_{i2} < 0.5\right) + \omega_i, & \text{if } A_i = 1 \end{cases},$$

where the ω_i's are independent $N(0, 1)$ random variables. The survival time T_i is generated by

$$T_i = 1 + X_{i1} + X_{i2} + A_i\left(3 - 3X_{i1} - 1.5X_{i2}\right) + \left[0.5 + A_i\left(1 + X_{i1} + X_{i2}\right)\right]\epsilon_i, \qquad (17)$$

where the ϵ_i's are independent normal random variables with mean zero and standard deviation 0.5. The observed response is $Y_i = \min\{T_i, C_i\}$. This configuration yields a 30% censoring rate. We consider estimating the τth quantile optimal IDR (τ=0.1 and 0.25)

τ	$\beta_{01}^{(\tau)}$	$\beta_{02}^{(\tau)}$	$\beta_{03}^{(\tau)}$	$Q_\tau(\boldsymbol{\beta})$
0.10	-1	0.896	-0.774	1.853
0.25	-1	1.140	-0.825	2.247

Table 4: Parameters indexing the quantile-optimal IDRs ($\tau = 0.1$ and 0.25) and the the maximally achievable τth quantile of the potential outcome (denoted by Q_τ) in $\mathbb{D}$ in Example 1, based on a Monte Carlo experiment ($n = 10^7$).

	$\tau = 0.1$			$\tau = 0.25$		
Method	$\beta_{02}^{(\tau)}$	$\beta_{03}^{(\tau)}$	Q_τ	$\beta_{02}^{(\tau)}$	$\beta_{03}^{(\tau)}$	Bias for Q_τ
Naive	-0.107(0.150)	0.007(0.295)	-0.169(0.065)	-0.177(0.13)	0.035(0.215)	-0.206(0.048)
New ($h_n = 0.08$)	-0.022(0.094)	0.020(0.162)	0.025(0.059)	-0.024(0.175)	-0.035(0.265)	-0.029(0.064)
New ($h_n = 0.10$)	-0.016(0.091)	0.029(0.165)	0.030(0.058)	0.019(0.200)	-0.055(0.294)	-0.031(0.067)
New ($h_n = 0.12$)	-0.009(0.086)	0.020(0.157)	0.027(0.058)	-0.014(0.195)	-0.046(0.277)	-0.008(0.065)
New ($h_n = 0.14$)	-0.020(0.094)	0.035(0.170)	0.038(0.054)	0.002(0.187)	-0.029(0.275)	0.003(0.069)

Table 5: Bias (with standard deviation in the parenthesis) of New and Naive for estimating $\boldsymbol{\beta}_0^{(\tau)}$ and Q_τ for Example 2.

within the class $\mathbb{D} = \left\{ \mathrm{I}\left(\beta_1 X_1 + \beta_2 + \beta_3 X_2 > 0\right) : |\beta_1| = 1, (\beta_2, \beta_3)^T \in \mathbb{R}^2 \right\}$. Similarly as for example 1, the parameters indexing the quantile-optimal IDRs ($\tau = 0.1$ and 0.25) and the the maximally achievable τth quantile of the potential outcome (denoted by Q_τ) in $\mathbb{D}$ were estimated based on a large Monte Carlo experiment with sample size $n = 10^7$ and treated as population parameter values, see Table 4.

To incorporate covariate-dependent censoring, we adopt the local Kaplan-Meier estimator (with bandwidth $h_n = 0.08, 0.1, 0.12, 0.14$) described in Remark 1 of Section 2 to estimate the propensity score. Table 5 summarizes the simulation results for New and Naive for $n = 500$ based on 300 replications. We observe that New has satisfactory performance and its performance is stable with respect to different choices of the bandwidth h. In contrast, Naive exhibits substantial bias for estimating $\boldsymbol{\beta}_{02}^{(\tau)}$ and Q_τ.

Example 3 (Two-stage dynamic individualized decision rule). The simulation setup is motivated by the example in Jiang et al. (2017a). The censoring time $C \sim \mathrm{Unif}\,(0, C_0)$, where C_0 is a positive constant. Let $s = 1$. Generate the data up to time s, $\left\{X_1, D_1, \mathsf{Z}^C = \mathrm{I}\big(\min\left(T_1, C\right) > s\big)\right\}$, from the following distributions: $X_1 \sim \mathrm{Unif}\,(0, 4)$, $D_1 \mid X_1 \sim \mathrm{Bernoulli}(0.5)$ and $T_1 \mid X_1, D_1 \sim \mathrm{Exp}\big(\lambda_1(X_1, D_1)\big)$ where $\lambda_1(\cdot)$ is a rate function to be specified later, and Z^C is an auxiliary variable that equals the product of $\dot{R}$ and Γ in

observed data model (13).

If $\mathsf{Z}^C = 1$, then the simulated patient is eligible for stage-two treatment. We generate the intermediate covariate X_2, second stage treatment D_2 and time T_2, representing the survival time after time s according to: $e \sim \text{Unif}\,(0,2)$, $X_2 \mid X_1, D_1 = 0.5X_1 - 0.4\big(D_1 - 0.5\big) + e$, $D_2 \mid X_1, D_1, X_2 \sim \text{Bernoulli}(0.5)$, $T_2 \mid X_1, D_1, X_2, D_2 \sim \text{Exp}\big(\lambda_2(X_1, D_1, X_2, D_2)\big)$, where $\lambda_2(\cdot)$ is a rate function to be specified later. For $\mathsf{Z}^C = 1$, the observed survival time and censoring status are $Y = \min\{(s+T_2), C\}$ and $\Delta = \text{I}\{(s+T_2) \leq C\}$, respectively; for $\mathsf{Z}^C = 0$, the observed survival time and censoring status are $Y = \min\{T_1, C\}$ and $\Delta = \text{I}\{T_1 \leq C\}$, respectively. Let $H_1 = \{X_1\}$ and $H_2 = \{X_1, D_1, X_2\}$. For rate functions for T_1 and T_2, consider three scenarios:

(a) $\lambda_1(H_1, D_1) = 0.5 \exp\big(1.75(D_1 - 0.5)(X_1 - 2)\big)$,
$\lambda_2(H_2, D_2) = 0.3 \exp\big(2.5(D_2 - 0.4)(X_2 - 2) - D_1(X_1 - 2)\big)$

(b) $\lambda_1(H_1, D_1) = 0.2 \exp\big(2(D_1 - 0.5)(-X_1 + 2)\big)$,
$\lambda_2(H_2, D_2) = 0.2 \exp\big(1.5(D_2 - 0.5)(X_2 - 2) + 0.3X_1 + 0.3X_2\big)$

(c) $\lambda_1(H_1, D_1) = 0.3 \exp\big(3(D_1 - 0.3)(X_1 - 3)\big)$,
$\lambda_2(H_2, D_2) = 0.3 \exp\big(2(D_2 - 0.5)(X_2 - 2) - 0.5(D_1 - 0.3)(X_1 - 3)\big)$

For each scenario, we consider two different choice of C_0 to achieve 15% and 40% overall censoring rate, respectively.

In this setup, T_1 serves as the underlying $\sum_{j\in\{1,2\}} \text{I}\{D_1 = A_j\}\dot{T}(A_j, \emptyset)$ in equation (12); T_2 is the survival time after initiation of the second stage treatment conditional on $\mathsf{Z}^C = 1$, hence $T_2 + s$ serves as $\sum_{j\in\{1,2\},\ k\in\{1,2\}} \text{I}\{D_1 = A_j, D_2 = B_k\}\dot{T}\big(A_j, B_k\big)$ in (12). By the interaction between D_2 and H_2 in $\lambda_2(\cdot)$ functions, the three scenarios share the same true optimal second-stage strategy for patients with $\mathsf{Z}^C = 1$, which is $d_2^{opt}(H_2) = \text{I}(-X_2 + 2 > 0)$. Suppose the class of IDRs is $\mathcal{D} = \big\{\boldsymbol{d_\xi} = (d_{1,\boldsymbol{\beta}}, d_{2,\boldsymbol{\zeta}}) : d_{1,\boldsymbol{\beta}}(X_1) = \text{I}\{\beta_1 X_1 + \beta_2 > 0\}, d_{2,\boldsymbol{\zeta}}(X_2) = \text{I}\{\zeta_1 X_2 + \zeta_2 > 0\}, |\beta_1| = 1, |\zeta_1| = 1\big\}$. Thus, $d_2^{opt}(H_2)$ is contained in $\mathcal{D}$, and the parameter indexing $d_2^{opt}(H_2)$ in $\mathcal{D}$ is $\zeta_{01} = -1, \zeta_{02} = 2$. However, for all three cases, the true value for $\boldsymbol{\beta}_0$ indexing the optimal first-stage treatment in $\mathcal{D}$ does not have close form representations. We used grid-search with sample size $n = 10^7$ for $\boldsymbol{\beta} \in \{-1\} \times [0, 4] \cup \{1\} \times [-4, 0]$ to obtain

Case	n	C%	Bias in β_{02}	Bias in ζ_{02}	Bias in $Q_{0.3}(\boldsymbol{\xi}_0)$
(a)	300	40	-0.052(0.499)	0.035(0.391)	0.525(0.445)
	500	40	-0.036(0.416)	0.031(0.339)	0.353(0.358)
	1000	40	-0.005(0.317)	0.043(0.282)	0.206(0.220)
	300	15	0.016(0.402)	-0.008(0.356)	0.331(0.345)
	500	15	0.010(0.329)	0.006(0.303)	0.247(0.262)
	1000	15	-0.008(0.255)	0.001(0.230)	0.139(0.187)
(b)	300	40	-0.030(0.648)	-0.023(0.371)	0.280(0.225)
	500	40	-0.031(0.581)	-0.015(0.344)	0.181(0.149)
	1000	40	-0.016(0.429)	-0.013(0.279)	0.120(0.105)
	300	15	-0.031(0.600)	0.005(0.332)	0.235(0.178)
	500	15	-0.037(0.485)	0.012(0.314)	0.155(0.137)
	1000	15	0.006(0.414)	-0.004(0.260)	0.101(0.090)
(c)	300	40	-0.084(0.433)	-0.009(0.342)	0.484(0.396)
	500	40	-0.056(0.372)	0.001(0.299)	0.380(0.347)
	1000	40	0.003(0.277)	0.008(0.250)	0.198(0.231)
	300	15	-0.074(0.404)	0.002(0.304)	0.379(0.329)
	500	15	-0.029(0.336)	-0.012(0.254)	0.261(0.286)
	1000	15	-0.015(0.249)	0.001(0.229)	0.150(0.167)

Table 6: Simulation Results about the *bias* (standard deviation of the estimates given in the parentheses) of $\widehat{\boldsymbol{\xi}}_n$ relative to $\boldsymbol{\xi}_0$ and of $\widehat{Q}_{0.3}(\widehat{\boldsymbol{\xi}})$ relative to $Q_{0.3}(\boldsymbol{\xi}_0)$. We used 400 replicates. The results for $\hat{\beta}_1$ and $\hat{\zeta}_1$ were omitted, since they ought to be from the set $\{-1, 1\}$ by the normalization method and our estimator correctly estimate them in each run. True value can be found in Table 7.

$\boldsymbol{\beta}_0$ in these cases, where the search space is the largest set of identifiable $\boldsymbol{\beta}$ since X_1 has support $[0, 4]$.

With the above setup, we generate a random sample $\{X_{i1}, D_{i1}, Z_i^C X_{i2}, Z_i^C D_{i2}, Y_i, \Delta_i\}$, $i = 1, \ldots, n$, where $n = 300, 500$ and 1000 are considered. We assume the randomization probability π_1 and π_2 are known, and applied the proposed method to estimate the optimal IDR with $\tau = 0.3$. Table 6 reports the simulation estimates of bias and standard deviations of the parameter indexing the quantile optimal dynamic regime $\widehat{\boldsymbol{\xi}}_n$. It also reports estimates of bias and standard deviation of the plug-in estimator of maximal achievable 0.3-quantile, $\widehat{Q}_{0.3}\{T^*(\boldsymbol{d}_{\widehat{\boldsymbol{\xi}}_n})\}$. We observe the proposed method reliably estimated the optimal two-stage IDR. The average biases and standard deviations of $\hat{\beta}_{02}$ and $\hat{\zeta}_{02}$ decrease as sample size increases. The lower censoring rate corresponds to better performance.

	$\boldsymbol{\beta}_0$	$\boldsymbol{\zeta}_0$	$Q_{0.3}(\boldsymbol{\xi}_0)$
Case (a)	$(-1, 2.00)^T$	$(-1, 2)^T$	1.524
Case (b)	$(1, -1.95)^T$	$(-1, 2)^T$	1.566
Case (c)	$(-1, 2.94)^T$	$(-1, 2)^T$	2.132

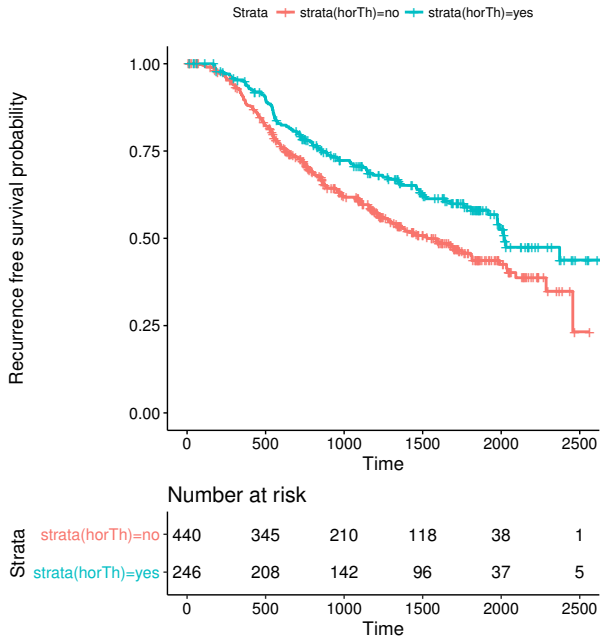
Table 7: True values of $\boldsymbol{\xi}_0 = (\boldsymbol{\beta}_0^T, \boldsymbol{\zeta}_0^T)^T$ indexing the 0.3 quantile-optimal DTR

6 Analysis of GBSG2 study data

To illustrate the proposed method, we analyze the data from the GBSG2 study conducted by the German Breast Cancer Study Group (Schumacher et al., 1994; Schmoor et al., 1996). The study investigated the efficacy of four combinations of treatments: three versus six cycles of chemotherapy with or without the adjuvant hormonal therapy with Tamoxifen. The outcome of interest is the recurrence-free survival time in days. The dataset, available in the R package `TH.data` (Hothorn, 2017), contains information on 686 patients of whom 56% had censored outcomes. The covariates include the age at diagnosis, the menopausal status, tumour size, tumour grade, the number of positive lymph nodes, estrogen receptor (ER) and progesterone receptor (PR) expression level in the tumour tissue. Earlier work on this study provided strong evidence that six cycles of chemotherapy is not superior to three cycles with respect to recurrence-free survival. Our analysis therefore focuses on IDRs regarding the assignment of adjuvant Tamoxifen therapy. Figure 2 depicts the estimated Kaplan-Meier curves for the groups of patients with and without Tamoxifen therapy.

First, we estimate the probability of receiving Tamoxifen. In this study, about two thirds of the recruited patients were randomized, and those who were not randomized chose the treatment by personal or professional preference. Because the randomization status is masked in the anonymous version of this data, we consider a working propensity score model by logistic regression. We observe that the randomization status is not significantly associated with the survival (Schmoor et al., 1996). Let A denote the hormonal therapy status ($A = 0$: did not receive; $A = 1$: received). We first fit a logistic regression model using A as the response and all available covariates. We then perform a best subset selection using the R package `bestglm` (McLeod and Xu, 2010), and obtained the model $\text{logit}\{\pi_A(\boldsymbol{X}; \boldsymbol{\gamma})\} =$

Figure 2: Plot of the Kaplan-Meier estimator of survival functions of T for $A = 0, 1$ respectively



$\gamma_0 + \gamma_1$MNST, where MNST is the binary menopausal status of patients. Using this selected model, we obtain the estimated propensity score 0.203 for premenopausal patients, and 0.472 for postmenopausal patients. The dependency of propensity score on MNST is mostly due to a modification on the protocol for randomization starting from the third year of GBSG2 recruitment. We also tried the propensity score model with all covariates and found that it leads to almost the same recommendations as measured by the match ratio (percentage of times two decision rules make the same treatment recommendations), which is above 98% for both of the two classes of regimes under consideration.

Motivated by the extensive work in the medical literature on Tamoxifen's molecular level mechanism and its clinical long-term effects, we consider IDRs that depend on the following three variables.

1. ER: The role of estrogen receptor expression as a predictive factor guiding the allocation of tamoxifen is well recognized. A large meta-analysis of randomized clinical trials demonstrated that high-ER patients respond better to Tamoxifen compared with low-ER patients (Group et al., 1998).

Regimes	β_2	β_3	β_4
$\mathcal{D}_1$	-1.23	0.94	-0.14
	(-3.39, -0.88)	(0.87, 2.02)	(-1.21, 2.39)
$\mathcal{D}_2$	-1.26	0.97	/
	(-2.41, -1.07)	(0.43, 2.01)	/

Table 8: Estimated parameters indexing the quartile-optimal IDR and 90% m-out-of-n bootstrap confidence intervals for the GBSG2 study

2. PR: Progesterone receptor expression is routinely measured for breast cancer patients as an important prognostic factor. However, its predictive power for the efficacy of Tamoxifen is still not well understood. It was observed that breast cancer patients with both high ER and high PR ('double positive') have the best chance of surviving (Bardou et al., 2003).

3. Age: Age is an important risk factor in breast cancer. We speculate that it may also contribute to how well patients respond to the adjuvant Tamoxifen therapy.

Because ER and PR are both highly skewed and have the minimal value 0 in this dataset, we adopt the transformation $\text{LER} = \log_{10}(\text{ER}+1)$ and $\text{LPR} = \log_{10}(\text{PR}+1)$. Age is linearly normalized to be between 0 and 1, and is denoted by NAGE.

We first consider the class of IDRs $\mathcal{D}_1$ that depend on all three variables:

$$\mathcal{D}_1 = \big\{\text{I}\big(\beta_1\text{LER} + \beta_2 + \beta_3\text{LPR} + \beta_4\text{NAGE} > 0\big) : \beta_1 = 1, \beta_2, \beta_3, \beta_4 \in \mathbb{R}\big\}.$$

We restrict the sign of ER to be positive based on evidence from the clinical practice (Hammond et al., 2010). We estimate the IDR in the class $\mathcal{D}_1$ that maximizes the first quartile ($\tau = 0.25$) of the recurrence-free survival time. We examine the dependence of the censoring time C on A and all seven prognostic factors by Cox regression and conclude that the independent censoring assumption is plausible. Furthermore, since GBSG2 has a high censoring rate and relatively short follow-up time, hereafter we use the artificial censoring technique (with M being set as 1550 days)in Remark 2 in Section 2.3 to improve stability of the proposed method.

The estimated parameter indexing the quartile-optimal IDR is $\widehat{\boldsymbol{\beta}}_{n,\mathcal{D}_1} = (1, -1.23, 0.94, -0.14)^T$. This regime leads to an estimated quartile survival time of $\widehat{Q}_{0.25}\left(\widehat{\boldsymbol{\beta}}_{n,\mathcal{D}_1}\right) = 1246$ days with

approximately 81.6% of patients being recommended to treatment. In contrast, the Kaplan-Meier estimator of the first quartile of the observed survival time is 727 (90% confidence interval = $(622, 805)$). The first row in Table 8 reported the 90% smoothed bootstrap confidence interval (Section 3.3) for each coefficient except β_1 based on 400 bootstrap samples. The coefficient of NAGE, β_4, is insignificant at the 0.1 level, which suggests that age may not be an important variable for determining Tamoxifen.

Next, we estimate quartile-optimal IDR in the following simplified class of IDRs to obtain a concise rule,

$$\mathcal{D}_2 = \{\mathrm{I}\left(\beta_1 \mathrm{LER} + \beta_2 + \beta_3 \mathrm{LPR} > 0\right) : \beta_1 = 1, \beta_2, \beta_3 \in \mathbb{R}\}.$$

The estimated parameter indexing the quartile-optimal IDR in $\mathcal{D}_2$ is $\widehat{\boldsymbol{\beta}}_{n,\mathcal{D}_2} = (1, -1.26, 0.97)^T$, which leads to an estimated quartile survival time of $\widehat{Q}_{0.25}\left(\widehat{\boldsymbol{\beta}}_{n,\mathcal{D}_2}\right) = 1246$ days with approximately 82.3% of patients being recommended to treatment. The second row of Table 8 reported the 90% smoothed bootstrap confidence intervals for β_2 and β_3. The coefficient for LPR is significant. Also, because the estimated $\hat{\beta}_3$ for LPR is about 1, we conclude that LPR is as important as the well-established predictive factor LER in developing an IDR that optimizes the first quartile survival time.

References

Abrevaya, J. and Huang, J. (2005). On the bootstrap of the maximum score estimator. *Econometrica*, 73(4):1175–1204.

Bai, X., Tsiatis, A. A., Lu, W., and Song, R. (2017). Optimal treatment regimes for survival endpoints using a locally-efficient doubly-robust estimator from a classification perspective. *Lifetime Data Analysis*, 23(4):585–604.

Banerjee, M., Durot, C., Sen, B., et al. (2019). Divide and conquer in nonstandard problems and the super-efficiency phenomenon. *The Annals of Statistics*, 47(2):720–757.

Banerjee, M. and McKeague, I. W. (2007). Confidence sets for split points in decision trees. *The Annals of Statistics*, 35(2):543–574.

Bardou, V.-J., Arpino, G., Elledge, R. M., Osborne, C. K., and Clark, G. M. (2003). Progesterone receptor status significantly improves outcome prediction over estrogen receptor

status alone for adjuvant endocrine therapy in two large breast cancer databases. *Journal of Clinical Oncology*, 21(10):1973–1979.

Bickel, P. J., Götze, F., and van Zwet, W. R. (2012). Resampling fewer than n observations: gains, losses, and remedies for losses. In *Selected works of Willem van Zwet*, pages 267–297. Springer.

Bickel, P. J. and Sakov, A. (2008). On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statistica Sinica*, 18:967–985.

Chakraborty, B., Laber, E. B., and Zhao, Y. (2013). Inference for optimal dynamic treatment regimes using an adaptive m-out-of-n bootstrap scheme. *Biometrics*, 69(3):714–723.

Chakraborty, B. and Moodie, E. E. (2013). *Statistical Methods for Dynamic Treatment Regimes: Reinforcement Learning, Causal Inference, and Personalized Medicine*. Springer Science & Business Media.

Chakraborty, B., Murphy, S., and Strecher, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research*, 19(3):317–343.

Cui, Y., Zhu, R., and Kosorok, M. (2017). Tree based weighted learning for estimating individualized treatment rules with censored data. *Electronic journal of statistics*, 11(2):3927.

Delgado, M. A., Rodrıguez-Poo, J. M., and Wolf, M. (2001). Subsampling inference in cube root asymptotics with an application to manski's maximum score estimator. *Economics Letters*, 73(2):241–250.

Delsol, L. and Van Keilegom, I. (2020). Semiparametric m-estimation with non-smooth criterion functions. *Annals of the Institute of Statistical Mathematics*, 72(2):577–605.

Díaz, I., Savenkov, O., and Ballman, K. (2018). Targeted learning ensembles for optimal individualized treatment rules with time-to-event outcomes. *Biometrika*, 105(3):723–738.

Goldberg, Y. and Kosorok, M. R. (2012). Q-learning with censored data. *Annals of Statistics*, 40(1):529.

Gonzalez-Manteiga, W. and Cadarso-Suarez, C. (1994). Asymptotic properties of a generalized kaplan-meier estimator with some applications. *Communications in Statistics-Theory and Methods*, 4(1):65–78.

Group, E. B. C. T. C. et al. (1998). Tamoxifen for early breast cancer: an overview of the randomised trials. *The Lancet*, 351(9114):1451–1467.

Hager, R., Tsiatis, A. A., and Davidian, M. (2018). Optimal two-stage dynamic treatment regimes from a classification perspective with censored survival data. *Biometrics*, 74(4):1180–1192.

Hammond, M. E. H., Hayes, D. F., Dowsett, M., Allred, D. C., Hagerty, K. L., Badve, S., Fitzgibbons, P. L., Francis, G., Goldstein, N. S., Hayes, M., et al. (2010). American society of clinical oncology/college of american pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Archives of Pathology & Laboratory Medicine*, 134(7):e48–e72.

Hothorn, T. (2017). *TH.data: TH's Data Archive*. `https://cran.r-project.org/web/packages/TH.data/`.

Jiang, R., Lu, W., Song, R., and Davidian, M. (2017a). On estimation of optimal treatment regimes for maximizing t-year survival probability. *Journal of the Royal Statistical Society: Series B*, 79(4):1165–1185.

Jiang, R., Lu, W., Song, R., Hudgens, M. G., Naprvavnik, S., et al. (2017b). Doubly robust estimation of optimal treatment regimes for survival data with application to an hiv/aids study. *The Annals of Applied Statistics*, 11(3):1763–1786.

Kim, J. K. and Pollard, D. (1990). Cube root asymptotics. *The Annals of Statistics*, 18:191–219.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York.

Kosorok, M. R. and Moodie, E. E. (2016). *Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine*. ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, ASA, Alexandria, VA.

Laber, E. and Zhao, Y. (2015). Tree-based methods for individualized treatment regimes. *Biometrika*, 102(3):501–514.

Lavori, P. W. and Dawson, R. (2000). A design for testing clinical strategies: biased adaptive within-subject randomization. *Journal of the Royal Statistical Society: Series A*, 163:29–38.

Léger, C. and MacGibbon, B. (2006). On the bootstrap in cube root asymptotics. *Canadian Journal of Statistics*, 34(1):29–44.

Luedtke, A. and van der Laan, M. J. (2016). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *The Annals of Statistics*, 44(2):713–742.

Matsouaka, R. A., Li, J., and Cai, T. (2014). Evaluating marker-guided treatment selection strategies. *Biometrics*, 70(3):489–499.

McLeod, A. and Xu, C. (2010). bestglm: Best subset glm. `https://cran.r-project.org/web/packages/bestglm/`.

Mebane Jr, W. R. and Sekhon, J. S. (2011). Genetic optimization using derivatives: the rgenoud package for r. *Journal of Statistical Software*, 42:1–26.

Mitchell, M. (1998). *An Introduction to Genetic Algorithms*. MIT press.

Moodie, E. E. and Richardson, T. S. (2010). Estimating optimal dynamic regimes: Correcting bias under the null. *Scandinavian Journal of Statistics*, 37(1):126–146.

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B*, 65(2):331–366.

Murphy, S. A. (2005a). An experimental design for the development of adaptive treatment strategies. *Statistics in medicine*, 24(10):1455–1481.

Murphy, S. A. (2005b). A generalization error for q-learning. *Journal of Machine Learning Research*, 6:1073–1097.

Murphy, S. A. (2008). An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine*, 24:1455–1481.

Neyman, J. (1923). Sur les applications de la théorie des probabilités aux experiences agricoles: Essai des principes. *Roczniki Nauk Rolniczych*, 10:1–51.

Orellana, L., R. A. and Robins, J. (2010). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part i: Main content. *The International Journal of Biostatistics*, 6.

Patra, R. K., Seijo, E., and Sen, B. (2018). A consistent bootstrap procedure for the maximum score estimator. *Journal of Econometrics*, 205(2):488–507.

Qian, M., Chakraborty, B., Maiti, R., and Cheung, Y. K. (2021). A sequential significance test for treatment by covariate interactions. *Statistica Sinica*. In press.

Qian, M. and Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *The Annals of Statistics*, 39(2):1180–1210.

Robins, J., Hernan, M., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11:550–560.

Robins, J. and Rotnitzky, A. G. (2014). Discussion of "dynamic treatment regimes: Technical challenges and applications". *Electronic Journal of Statistics*, 8:1273–1289.

Robins, J.M., O. L. and Rotnitzky, A. (2008). Estimation and extrapolation of optimal treatment and testing strategies. *Statistics in Medicine*, 27:4678–4721.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, 6(1):34–58.

Rubin, D. B. (1986). Which ifs have causal answers. *Journal of the American Statistical Association*, 81:961–962.

Schmoor, C., Olschewski, M., and Schumacher, M. (1996). Randomized and non-randomized patients in clinical trials: experiences with comprehensive cohort studies. *Statistics in Medicine*, 15(3):263–271.

Schumacher, M., Bastert, G., Bojar, H., Huebner, K., Olschewski, M., Sauerbrei, W., Schmoor, C., Beyerle, C., Neumann, R., and Rauschecker, H. (1994). Randomized 2 x 2 trial evaluating hormonal treatment and the duration of chemotherapy in node-positive breast cancer patients. german breast cancer study group. *Journal of Clinical Oncology*, 12(10):2086–2093.

Sen, B., Banerjee, M., Woodroofe, M., et al. (2010). Inconsistency of bootstrap: The grenander estimator. *The Annals of Statistics*, 38(4):1953–1977.

Shi, C., Lu, W., and Song, R. (2018). A massive data framework for m-estimators with cubic-rate. *Journal of the American Statistical Association*, 113(524):1698–1709.

Simoneau, G., Moodie, E. E., Nijjar, J. S., Platt, R. W., Investigators, S. E. R. A. I. C., et al. (2020). Estimating optimal dynamic treatment regimes with survival outcomes. *Journal of the American Statistical Association*, 115:1531–1539.

Song, R., Wang, W., Zeng, D., and Kosorok, M. (2015). Penalized q-learning for dynamic treatment regimens. *Statistica Sinica*, 25:901–920.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer-Verlag, New York.

Wahed, A. S. (2009). Estimation of survival quantiles in two-stage randomization designs. *Journal of Statistical Planning and Inference*, 139(6):2064–2075.

Wang, L., Zhou, Y., Song, R., and Sherwood, B. (2018). Quantile-optimal treatment regimes. *Journal of the American Statistical Association*, 113(523):1243–1254.

Watkins, C. and Dayan, P. (1992). Q-learning. *Machine Learning*, 8:279–292.

Wu, Y. and Wang, L. (2021). Resampling-based confidence intervals for model-free robust inference on optimal treatment regimes. *Biometrics*, 77(2):465–476.

Xu, Y., Müller, P., Wahed, A. S., and Thall, P. F. (2016). Bayesian nonparametric estimation for dynamic treatment regimes with sequential transition times. *Journal of the American Statistical Association*, 111(515):921–950.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics*, 68(4):1010–1018.

Zhang, Y., Laber, E., Davidian, M., and Tsiatis, A. (2018). Estimation of optimal treatment regimes using lists. 113:1541–1549.

Zhao, Y., Zeng, D., Laber, E. B., Song, R., Yuan, M., and Kosorok, M. R. (2015a). Doubly robust learning for estimating individualized treatment with censored data. *Biometrika*, 102(1):151–168.

Zhao, Y., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499):1106–1118.

Zhao, Y. Q., Zeng, D., Laber, E. B., and Kosorok, M. R. (2015b). New statistical learning methods for estimating optimal dynamic treatment regimes. *Journal of the American Statistical Association*, 110:583–598.

Zhou, L. (2006). A simple censored median regression estimator. *Statistica Sinica*, 16(3):1043–1058.

Zhu, R., Zhao, Y.-Q., Chen, G., Ma, S., and Zhao, H. (2017). Greedy outcome weighted tree learning of optimal personalized treatment rules. *Biometrics*, 73(2):391–400.

Appendix: Regularity Conditions

We introduce below a set of regularity conditions needed to establish the statistical theory. The proof of the theory is given in the online supplement.

C1 Let L denote the end of the study. The censoring variable C has a continuously differentiable density function which is bounded away from infinity on $\big(0, L\big)$. There exists a constant $\eta > 0$ such that $G_C(L) > \eta > 0$. The densities $f_0(t|\boldsymbol{X})$ and $f_1(t|\boldsymbol{X})$ are uniformly bounded away from infinity, almost surely in $\boldsymbol{X}$; and $\sup_{\boldsymbol{\beta}\in\mathbb{B}^o} Q_\tau\big\{T^*\big(d_{\boldsymbol{\beta}}\big)\big\} < L$. There exist positive constants κ_1 and δ, such that $\inf_{\boldsymbol{\beta}\in\mathbb{B}^o} \inf_{|m-m_0|\le\delta} f_{T^*\left(d_{\boldsymbol{\beta}}\right)}(m) \ge \kappa_1$, where $m_0 = \sup_{\boldsymbol{\beta}\in\mathbb{B}^o} Q_\tau\big\{T^*\big(d_{\boldsymbol{\beta}}\big)\big\}$.

C2 The probability density function of X_1 conditional on $\tilde{\boldsymbol{X}}$ is continuously differentiable. The angular component of $\boldsymbol{X}$, considered as a random element of the sphere $\mathbb{S} \in \mathbb{R}^p$, has a bounded and continuous density.

C3 The population parameter $\boldsymbol{\beta}_0 = (\beta_{01}, \widetilde{\boldsymbol{\beta}}^T)^T$ indexing the optimal IDR is unique in $\mathbb{B}^o$.

C4 The $(p-1)\times(p-1)$ matrix $\Lambda(\widetilde{\boldsymbol{\beta}}, h)|_{\widetilde{\boldsymbol{\beta}}=\widetilde{\boldsymbol{\beta}}_0, h=h_0}$, defined in the proof of Lemma 1 in the online supplement, is negative definite.

Remark. Condition (C1) is common in survival analysis, where L is the maximum follow-up time. The survival time is not observed if it exceeds L. Condition (C2) has to do with

population parameter identifiability, as discussed in Section 2.2. We assume after a possible rearranging of elements in $\boldsymbol{X}$, the density of X_1 conditional on $\tilde{\boldsymbol{X}}$ is continuously differentiable for every $\tilde{\boldsymbol{X}}$ almost surely. If all covariates are discrete, the problem of estimating an optimal IDR actually becomes simpler in some sense as there are finite many decision rules. One can directly compare the estimated value functions. Condition (C3) is standard for index models. Condition (C4) is needed for evaluating the Hessian matrix when establishing the limiting distribution of $\widehat{\boldsymbol{\beta}}_n$.

Supplement to "Transformation-Invariant Learning of Optimal Individualized Decision Rules with Time-to-Event Outcomes"

Yu Zhou, Lan Wang, Rui Song and Tuoyi Zhao

1 Proof of theory

We first prove Theorem 2, which is useful for studying the properties of $\widetilde{\boldsymbol{\beta}}_n$.

Proof of Theorem 2. Write $m_0 = V_{opt}$ and $\widehat{m}_n = \widehat{V}_n$. Recall that as discussed in Section 3.2, motivated by the first-order-optimization condition of the estimator $\widehat{Q}_\tau(\boldsymbol{\beta}, \widehat{G}_C)$, we have the following alternative expression of m_0 and $\hat{m}_n$:

$$
\begin{aligned}
&g\big(\cdot, \boldsymbol{\beta}, m, G\big) = \frac{R\big(\boldsymbol{\beta}\big)}{0.5 G_C\big(Y\big)} I\big(Y - m > 0\big), \\
&m_0 = \sup\big\{m : \sup_{\boldsymbol{\beta} \in \mathbb{B}^o} Pg\big(\cdot, \boldsymbol{\beta}, m, G_C\big) \geq 1 - \tau\big\}, \\
&\widehat{m}_n = \sup\big\{m : \sup_{\boldsymbol{\beta} \in \mathbb{B}^o} Pg\big(\cdot, \boldsymbol{\beta}, m, \widehat{G}_C\big) \geq 1 - \tau\big\}.
\end{aligned}
$$

By the permanence property of Donsker Class (e.g. Example 2.10.8 in Van Der Vaart and

[1]Yu Zhou is a machine learning engineer at Roku. Email: izhou@roku.com. Lan Wang is Professor, Department of Management Science, University of Miami. Email: lanwang@mbs.miami.edu. Rui Song is Professor, Department of Statistics, North Carolina State University. Email: rsong@ncsu.edu. Tuoyi Zhao is a Ph.D. student, Department of Management Science, University of Miami. Wang, Zhou and Zhao's research was supported by NSF FRGMS-1952373. Song's research was supported by NSF DMS-2113637.

Wellner (1996)), the class of functions

$$\mathcal{F} := \Big\{ g\big(\cdot, \boldsymbol{\beta}, m, G\big) = \Delta \frac{I\big\{A = d_{\boldsymbol{\beta}}\big(\boldsymbol{X}\big)\big\}}{0.5 G\big(Y\big)} I\big(Y - m > 0\big) : \boldsymbol{\beta} \in \mathbb{B}^o, m \in [0, L), \quad (1)$$
$$G\big(\,\cdot\,\big) \text{ is a positive decreasing function uniformly lower bounded by } \eta \text{ over } [0, L) \Big\}$$

is *Donsker*, since the set of regimes $d_{\boldsymbol{\beta}}\big(\,\cdot\,\big)$ and the class of functions $\big\{I\big(Y - m > 0\big) : m \in [0, L)\big\}$ are both VC-subgraph classes, and $G^{-1}\big(Y\big)$ is uniformly bounded in $\mathcal{F}$. Consider the empirical process indexed by functions from $\mathcal{F}$: $\mathbb{G}_n = \sqrt{n}\big(\mathbb{P}_n - P\big)$, where $Pg\big(Z, \boldsymbol{\beta}, m, G\big) = \int g\big(z, \boldsymbol{\beta}, m, G\big) P\big(dz\big)$ takes expectation over the distribution of $Z := \{X, A, Y, \Delta\}$; and $\mathbb{P}_n g\big(Z, \boldsymbol{\beta}, m, G\big) = n^{-1} \sum_{i=1}^n g\big(Z_i, \boldsymbol{\beta}_0, m, G\big)$. We would like to stress that we use the notation $Pg\big(Z, \boldsymbol{\beta}, m, \widehat{G}_C\big)$ to mean $Pg\big(Z, \boldsymbol{\beta}, m, G\big)$ being evaluated at $G = \widehat{G}_C$ and hence it is a random quantity.

We first consider perturb m_0 with a small $\epsilon > 0$. Let $\hat{\varsigma} = \big(m_0 + \epsilon, \widehat{G}_C\big)$, $\varsigma_0 = \big(m_0 + \epsilon, G_C\big)$. For any given $\boldsymbol{\beta}$ and G, the function $g\big(\cdot, \boldsymbol{\beta}, m, G\big)$ is decreasing in m. Note that $Pg\big(\cdot, \boldsymbol{\beta}, m, G\big) = 1 - F_{T^*(d_{\boldsymbol{\beta}})}(m)$, where $F_{T^*(d_{\boldsymbol{\beta}})}(\cdot)$ denotes the marginal cumulative function of the potential survival time $T^*\big(d_{\boldsymbol{\beta}}(\boldsymbol{X})\big)$. Observing that $\sup_{\boldsymbol{\beta} \in \mathbb{B}^o} Pg(\cdot, \boldsymbol{\beta}, m_0, G_C) = 1 - \tau$. By Taylor expansion and condition (C1), $\exists \kappa_1 > 0$ such that for all $\epsilon > 0$, $\sup_{\boldsymbol{\beta} \in \mathbb{B}^o} Pg(\cdot, \boldsymbol{\beta}, \varsigma_0) < 1 - \tau - \epsilon\kappa_1$. We have

$$\sup_{\boldsymbol{\beta} \in \mathbb{B}^o} \mathbb{P}_n g\big(\cdot, \boldsymbol{\beta}, \hat{\varsigma}\big) < 1 - \tau - \epsilon\kappa_1 + U_{n,1} + \sup_{\boldsymbol{\beta} \in \mathbb{B}^o} \big|(\mathbb{P}_n - P) g\big(\cdot, \boldsymbol{\beta}, \hat{\varsigma}\big)\big|, \quad (2)$$

where $U_{n,1} = \sup_{\boldsymbol{\beta} \in \mathbb{B}^o} \big|P\big\{g\big(\cdot, \boldsymbol{\beta}, \varsigma_0\big) - g\big(\cdot, \boldsymbol{\beta}, \hat{\varsigma}\big)\big\}\big|$. Note that

$$\begin{aligned} & P\big\{g\big(\cdot, \boldsymbol{\beta}, \varsigma_0\big) - g\big(\cdot, \boldsymbol{\beta}, \hat{\varsigma}\big)\big\} \\ = \;& P\Big\{2\Delta\big(A d_{\boldsymbol{\beta}}\big(\boldsymbol{X}\big) + (1 - A)(1 - d_{\boldsymbol{\beta}}\big(\boldsymbol{X}\big))\big)\big(G_C^{-1}(Y) - \widehat{G}_C^{-1}(Y)\big) I\big(Y - m_0 - \epsilon > 0\big)\Big\} \\ \leq \;& 2\eta^2 \int \big|\widehat{G}_C(Y) - G_C(Y)\big| P dz \end{aligned} \quad (3)$$

It follows from Csörgő and Horváth (1983) that $U_{n,1} = o\big(n^{-1/2+\gamma_0}\big)$, almost surely, for any $\gamma_0 > 0$.

Since $\mathcal{F}$ defined in (1) is Donsker,

$$\sup_{\boldsymbol{\beta}\in\mathbb{B}^o,m} \left|\left(\mathbb{P}_n - P\right)\, g\big(\cdot,\boldsymbol{\beta},m,\widehat{G}_C\big)\right| = O_p\big(n^{-1/2}\big). \tag{4}$$

Denote the left hand side of (4) by $U_{n,2}$.

Let the ϵ in (2) take value $\epsilon_{1,n} = \big(U_{n,1}+U_{n,2}\big)/\kappa_1$, which is a sequence of positive numbers that converges to zero. Then for all n large enough

$$\sup_{\boldsymbol{\beta}} \mathbb{P}_n g\big(\cdot,\boldsymbol{\beta}\in\mathbb{B}^o, m_0+\epsilon_{1,n},\widehat{G}_C\big) < 1-\tau-\epsilon_{1,n}\kappa_1+U_{n,1}+U_{n,2} = 1-\tau.$$

Hence, by the monotonicity of g in m, we have $\widehat{m}_n \le m_0+\epsilon_{1,n}$ for all n sufficiently large, which implies that $\widehat{m}_n \le m_0 + o_p\big(n^{-1/2+\gamma_0}\big)$ for an arbitrarily small $\gamma_0 > 0$.

In the other direction, there exists a positive constant κ_2 such that $\sup_{\boldsymbol{\beta}\in\mathbb{B}^o} Pg\big(\cdot,\boldsymbol{\beta},m_0-\epsilon,G_C\big) > 1-\tau+\epsilon\kappa_2$ for all small enough $\epsilon > 0$. Similarly, we can verify that there exists a sequence of positive numbers $\epsilon_{2,n} = o_p\big(n^{-1/2+\gamma_0}\big)$ such that $\widehat{m}_n \ge m_0-\epsilon_{2,n}$ for all $\gamma_0 > 0$. Combining the stochastic lower and upper bounds for $\hat{m}_n$, we conclude that $\widehat{m}_n = m_0 + o_p\big(n^{-1/2+\gamma_0}\big)$ for all $\gamma_0 > 0$. □

To derive the limit of $n^{1/3}(\widetilde{\boldsymbol{\beta}}_n - \widetilde{\boldsymbol{\beta}}_0)$ stated in Theorem 1, we next prove a preliminary result Lemma 1, which establishes that $\widetilde{\boldsymbol{\beta}}_n$ converges to $\widetilde{\boldsymbol{\beta}}_0$ at the rate $n^{-1/3}$.

Lemma 1. *Suppose conditions (C1)-(C4) are satisfied, then*

$$\|\widetilde{\boldsymbol{\beta}}_n - \widetilde{\boldsymbol{\beta}}_0\| = O_p(n^{1/3}).$$

Proof of Lemma 1. We first prove that $\|\widetilde{\boldsymbol{\beta}}_n - \widetilde{\boldsymbol{\beta}}_0\| = o_P\big(1\big)$, where $\|\cdot\|$ stands for the Euclidean norm. We note that by the definition of m_0 and $\widehat{m}_n$ in the proof of Theorem 2, we obtain an alternative expression of $\boldsymbol{\beta}_0$

$$\boldsymbol{\beta}_0 = \arg\max_{\boldsymbol{\beta}\in\mathbb{B}^o} Pg(\cdot,\boldsymbol{\beta},m_0,G_C).$$

That is, $\boldsymbol{\beta}_0$ is the parameter indexing the IDR that achieves the optimal value m_0. This

naturally leads to an alternative representation of $\widehat{\boldsymbol{\beta}}_n$, given by

$$\widehat{\boldsymbol{\beta}}_n = \underset{\boldsymbol{\beta}\in\mathbb{B}^o}{\arg\max}\; n^{-1}\sum_{i=1}^{n} g\left(\cdot, \boldsymbol{\beta}, \widehat{m}_n, \widehat{G}_C\right).$$

We will derive the consistency of $\widehat{\boldsymbol{\beta}}_n$ by checking the high-level sufficient conditions (A1)-(A5) in Theorem 1 in Delsol and Van Keilegom (2020), abbreviated as DVK in the sequel. Let $\mathcal{H}$ be the Cartesian product of $\mathbb{R}$ and $\mathcal{G}$, where $\mathcal{G}$ is the class of decreasing functions bounded between η and 1. Let $h = \big(m, G\big)$ be an arbitrary element in $\mathcal{H}$, let $h_0 = \big(m_0, G_C\big)$ denote the true unknown nuisance parameters and let $\widehat{h} = \big(\widehat{m}_n, \widehat{G}_C\big)$ be the vector of estimated nuisance parameters. In the following, we will use the supremum metric on $\mathcal{H}$, given by $d_{\mathcal{H}}\big(h, h_0\big) = \max\big\{|m - m_0|, \sup_t \big|G(t) - G_C(t)\big|\big\}$. Conditions (A1) and (A2) are trivially satisfied. Condition (A4) is also satisfied for the proposed estimator because the class of functions $\mathcal{F}$ defined in 1 is also *Donsker*. To check (A3) in DVK, we first note that the Kaplan-Meier estimator $\widehat{G}_C$ is uniformly consistent (Csörgő and Horváth (1983)). In addition, Theorem 2 verified that $\widehat{m}_n$ is consistent for m_0. Hence, $d_{\mathcal{H}}(\widehat{h}, h_0) \xrightarrow{P^*} 0$, and (A3) is satisfied. Finally, similarly as in (3), it is straightforward to check $\lim_{d_{\mathcal{H}}(h,h_0)\to 0} \sup_{\boldsymbol{\beta}} |Pg(\cdot, \boldsymbol{\beta}, h) - Pg(\cdot, \boldsymbol{\beta}, h_0)| = 0$, and (A5) in DVK holds. Thus, all conditions of DVK's Theorem 1 hold. This verifies the consistency of $\widehat{\boldsymbol{\beta}}_n$.

Under the normalization $|\beta_1| = 1$, we know $\widehat{\beta}_{n1} \xrightarrow{a.s.} \beta_{01}$, as $n \to \infty$. In the following, we will derive the convergence rate of $\widetilde{\boldsymbol{\beta}}_n$. For brevity, we only give detailed proof of the convergence rate of $\tilde{\boldsymbol{\beta}}_n$ under the assumption $\widehat{\beta}_{n1} = \beta_{01} = \beta_1 = 1$, and critical result for the case $\hat{\beta}_{n1} = \beta_{01} = \beta_1 = -1$ is given in the footnote.

For proving the $n^{-1/3}$ convergence rate of $\widetilde{\boldsymbol{\beta}}_n$, we will verify the high-level conditions (B1)-(B4) of Theorem 2 in DVK (2015). We have already verified $\|\widetilde{\boldsymbol{\beta}}_n - \widetilde{\boldsymbol{\beta}}_0\| = o_P\big(1\big)$. By the property of the Kaplan-Meier estimator (Csörgő and Horváth (1983)), we have $d_{\mathcal{H}}(\widehat{h}, h_0) = o_p\big(n^{-1/2+\gamma_0}\big)$ for an arbitrarily small $\gamma_0 > 0$. For an arbitrary $\gamma_0 > 0$, set $\nu_n = n^{1/2-\gamma_0}$, then $\nu_n d_{\mathcal{H}}\big(\widehat{h}, h_0\big) = O_{P^*}\big(1\big)$. Therefore, Condition (B1) is satisfied. Condition (B4) is fulfilled automatically by the construction of the estimator with $r_n = n^{1/3}$ and $\Phi_n(\delta) = \sqrt{\delta}$. In the following, we will check Conditions (B2) and (B3).

To verify (B2), we consider the following class of functions:

$$\mathcal{F}_{\delta,\delta_1'} = \big\{ g\big(\cdot, \boldsymbol{\beta}, h\big) - g\big(\cdot, \boldsymbol{\beta}_0, h\big) : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \le \delta, d_{\mathcal{H}}\big(h, h_0\big) \le \delta_1', |\beta_1| = 1,\ \widetilde{\boldsymbol{\beta}} \in \widetilde{\mathbb{B}} \big\}.$$

We will first define and demonstrate a square integrable envelope function $M_{\delta,\delta_1'}\big(Z\big)$ for the class $F_{\delta,\delta_1'}$.

$$\begin{aligned}
&\sup_{\boldsymbol{\beta},h} \big\{ \big| g\big(Z, \boldsymbol{\beta}, h\big) - g\big(Z, \boldsymbol{\beta}_0, h\big) \big| : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \le \delta,\ d_{\mathcal{H}}\big(h, h_0\big) \le \delta_1' \big\} \\
=&\sup_{\boldsymbol{\beta},h} \big\{ 2\Delta G^{-1}(Y) A \big| d_{\boldsymbol{\beta}}(\boldsymbol{X}) - d_{\boldsymbol{\beta}_0}(\boldsymbol{X}) \big| I\big(y - m < 0\big) : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \le \delta,\ d_{\mathcal{H}}\big(h, h_0\big) \le \delta_1' \big\} \\
\le&\sup_{\boldsymbol{\beta},h} \big\{ 2\eta^{-1} I\big(d_{\boldsymbol{\beta}}(\boldsymbol{X}) \ne d_{\boldsymbol{\beta}_0}(\boldsymbol{X})\big) : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \le \delta,\ d_{\mathcal{H}}\big(h, h_0\big) \le \delta_1' \big\} \\
\le&2\eta^{-1} \big\{ I\big[\delta \|\boldsymbol{X}\| > \boldsymbol{X}^T \boldsymbol{\beta}_0 > 0\big] + I\big[\delta \|\boldsymbol{X}\| > -\boldsymbol{X}^T \boldsymbol{\beta}_0 \ge 0\big] \big\} \\
\le&2\eta^{-1} I\Big[\frac{\delta}{\|\boldsymbol{\beta}_0\|} > \frac{\big|\boldsymbol{X}^T \boldsymbol{\beta}_0\big|}{\|\boldsymbol{X}\| \|\boldsymbol{\beta}_0\|} \Big] =: M_{\delta,\delta_1'}\big(Z\big). \qquad (5)
\end{aligned}$$

Clearly, $E\big[M_{\delta,\delta_1'}^2\big(Z\big)\big] = 4\eta^{-2} P\Big\{ \frac{\delta}{\|\boldsymbol{\beta}_0\|} > \big| \frac{\boldsymbol{X}^T}{\|\boldsymbol{X}\|} \frac{\boldsymbol{\beta}_0}{\|\boldsymbol{\beta}_0\|} \big| \Big\}$, where we need to evaluate the probability of the dot product between $\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|}$ and $\frac{\boldsymbol{\beta}_0}{\|\boldsymbol{\beta}_0\|}$ being between $(-\frac{\delta}{\|\boldsymbol{\beta}_0\|}, \frac{\delta}{\|\boldsymbol{\beta}_0\|})$. Let $\Phi_n\big(\delta\big) = \sqrt{\delta}$. By Condition (C2) in Subsection 3.1 on the density of the angular component of $\boldsymbol{X}$, $\exists\ \delta_0 > 0, \exists\ K_0 > 0$, such that $\forall\ \delta < \delta_0$,

$$E\big[M_{\delta,\delta_1'}^2\big(Z\big)\big] = 4\eta^{-2} E\Big\{ I\Big[\frac{\delta}{\|\boldsymbol{\beta}_0\|} > \Big| \frac{\boldsymbol{X}^T}{\|\boldsymbol{X}\|} \frac{\boldsymbol{\beta}_0}{\|\boldsymbol{\beta}_0\|} \Big| \Big] \Big\} \le K_0^2 \Phi_n^2(\delta). \qquad (6)$$

To verify Condition (B2) in DVK, it is sufficient to show there exists a positive constant K_1 such that for all $\delta \le \delta_0$,

$$\begin{aligned}
&E\big[\sup_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| < \delta,\ d_{\mathcal{H}}\left(h, h_0\right) \le \delta_1'} \big| \mathbb{P}_n g\big(\cdot, \boldsymbol{\beta}, h\big) - \mathbb{P}_n g\big(\cdot, \boldsymbol{\beta}_0, h\big) - P g\big(\cdot, \boldsymbol{\beta}, h\big) + P g\big(\cdot, \boldsymbol{\beta}_0, h\big) \big| \big] \\
\le\ & K_1 n^{-1/2} \sqrt{E\big[M_{\delta,\delta_1'}^2\big]}, \qquad (7)
\end{aligned}$$

Notice that $\mathcal{F}_{\delta,\delta_1'}$ is a VC-subgraph class. By Theorem 2.6.7 of Van Der Vaart and Wellner

(1996), $\mathcal{F}_{\delta,\delta_1'}$ satisfies the uniform entropy condition (Van Der Vaart and Wellner, 1996) and

$$\sup_{\delta<\delta_0,V}\int_0^1\sqrt{1+\log N\big(\epsilon\|M_{\delta,\delta_1'}\|_{L_2(V)},\mathcal{F}_{\delta,\delta_1'},L_2(V)\big)}d\epsilon<\infty, \tag{8}$$

where the supremum on distribution V takes over all discrete probability measure such that $\|M_{\delta,\delta_1'}\|_{L_2(V)}>0$. By Van Der Vaart and Wellner (1996) page 244, (8) implies (7) and hence Condition (B2) hold.

To verify Condition (B3), it is sufficient to prove the set of conditions in Remark 2(v) of DVK. We need to calculate the gradient and Hessian matrix of $Pg\big(\cdot,\boldsymbol{\beta},h\big)$ with respect to $\widetilde{\boldsymbol{\beta}}$. Note that we assumed $\beta_{01}=1$ in this main proof (the case $\beta_{01}=-1$ is only slightly different). We have

$$\begin{aligned}
& Pg\big(\cdot,\boldsymbol{\beta},h\big)\\
= \;& E\big\{d_{\boldsymbol{\beta}}\big(\boldsymbol{X}\big)\Delta G^{-1}\big(Y\big)I\big(Y>m\big)|A=1\big\}+E\big\{\big(1-d_{\boldsymbol{\beta}}\big(\boldsymbol{X}\big)\big)\Delta G^{-1}\big(Y\big)I\big(Y>m\big)|A=0\big\}\\
= \;& E_{\boldsymbol{X}}\big[I(\boldsymbol{X}^T\boldsymbol{\beta}>0)R_1\big(m,G|\boldsymbol{X}\big)\big]+E_{\boldsymbol{X}}\big[I(\boldsymbol{X}^T\boldsymbol{\beta}\le 0)R_0\big(m,G|\boldsymbol{X}\big)\big],
\end{aligned}$$

by the iterative expectation formula; where

$$\begin{aligned}
R_1\big(m,G|\boldsymbol{X}\big)&=E_C\Big[I\big(C>m\big)\int_m^C G^{-1}(u)f_1\big(u|\boldsymbol{X}\big)du\big|\boldsymbol{X}\Big],\\
R_0\big(m,G|\boldsymbol{X}\big)&=E_C\Big[I\big(C>m\big)\int_m^C G^{-1}(u)f_0\big(u|\boldsymbol{X}\big)du\big|\boldsymbol{X}\Big],
\end{aligned}$$

with $f_1(\cdot|\boldsymbol{X})$ being the conditional density function of $T^*\big(1\big)$ given $\boldsymbol{X}$ and $f_0(\cdot|\boldsymbol{X})$ being the conditional density function of $T^*\big(0\big)$ given $\boldsymbol{X}$. Define $q\big(h,\boldsymbol{X}\big)=R_1\big(m,G|\boldsymbol{X}\big)-R_0\big(m,G|\boldsymbol{X}\big)$, then $Pg\big(\cdot,\boldsymbol{\beta},h\big)=E_{\boldsymbol{X}}\big\{R_0\big(h|\boldsymbol{X}\big)+q\big(h,\boldsymbol{X}\big)I(\boldsymbol{X}^T\boldsymbol{\beta}>0)\big\}$. Under the restriction $\beta_1=1$, let $f_{\boldsymbol{X}_1|\tilde{\boldsymbol{X}}}(t)$ be the conditional density of $\boldsymbol{X}_1$ given $\tilde{\boldsymbol{X}}$. The gradient of

$Pg\big(\cdot,\boldsymbol{\beta},h\big)$ with respect to $\widetilde{\boldsymbol{\beta}}$ is

$$\begin{aligned}\Gamma(\widetilde{\boldsymbol{\beta}},h) =&\frac{\partial}{\partial\widetilde{\boldsymbol{\beta}}}E_{\boldsymbol{X}}\big[q\big(h,\boldsymbol{X}\big)I(\boldsymbol{X}_1>-\tilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{\beta}})\big]\\ =&\frac{\partial}{\partial\widetilde{\boldsymbol{\beta}}}E_{\tilde{\boldsymbol{X}}}\Big\{\int_{-\tilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{\beta}}}^{\infty}f_{\boldsymbol{X}_1|\tilde{\boldsymbol{X}}}(t)q\big(h,\big(t,\tilde{\boldsymbol{X}}^T\big)^T\big)dt\Big\}\\ =&E_{\tilde{\boldsymbol{X}}}\big\{\tilde{\boldsymbol{X}}f_{\boldsymbol{X}_1|\tilde{\boldsymbol{X}}}(-\tilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{\beta}})q\big(h,\big(-\tilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{\beta}},\tilde{\boldsymbol{X}}^T\big)^T\big)\big\}.\end{aligned}$$

Hence,

$$\Gamma\big(\widetilde{\boldsymbol{\beta}},h\big)|_{\widetilde{\boldsymbol{\beta}}=\widetilde{\boldsymbol{\beta}}_0}=E_{\tilde{\boldsymbol{X}}}\big\{\tilde{\boldsymbol{X}}f_{\boldsymbol{X}_1|\tilde{\boldsymbol{X}}}(-\tilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{\beta}}_0)q\big(h,\big(-\tilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{\beta}}_0,\tilde{\boldsymbol{X}}^T\big)^T\big)\big\}. \tag{9}$$

The Hessian matrix of $Pg\big(\cdot,\boldsymbol{\beta},h\big)$ with respect to $\widetilde{\boldsymbol{\beta}}$ is

$$\begin{aligned}\Lambda(\widetilde{\boldsymbol{\beta}},h) =&\frac{\partial}{\partial\widetilde{\boldsymbol{\beta}}}E_{\tilde{\boldsymbol{X}}}\Big\{\tilde{\boldsymbol{X}}f_{\boldsymbol{X}_1|\tilde{\boldsymbol{X}}}(-\tilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{\beta}})q\big(h,\big(-\tilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{\beta}},\tilde{\boldsymbol{X}}^T\big)^T\big)\Big\}\\ =&E_{\tilde{\boldsymbol{X}}}\Big[-\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^T\Big\{\frac{\partial}{\partial t}f_{\boldsymbol{X}_1|\tilde{\boldsymbol{X}}}(t)\big|_{t=-\tilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{\beta}}}\Big\}q\big(h,\big(-\tilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{\beta}},\tilde{\boldsymbol{X}}^T\big)^T\big)\\ &-\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^Tf_{\boldsymbol{X}_1\big|\tilde{\boldsymbol{X}}}\big(-\tilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{\beta}}\big)\Big\{\frac{\partial}{\partial t}q\big(h,\big(t,\tilde{\boldsymbol{X}}^T\big)^T\big)|_{t=-\tilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{\beta}}}\Big\}\Big].\end{aligned}$$

This Hessian matrix evaluated at $h = h_0$ is [1]:

$$
\begin{aligned}
&\Lambda(\tilde{\boldsymbol{\beta}}, h)|_{\tilde{\boldsymbol{\beta}}=\tilde{\boldsymbol{\beta}}_0, h=h_0} \\
=&E_{\tilde{\boldsymbol{X}}}\big(- \tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^T\big[\big\{\frac{\partial}{\partial t} f_{\boldsymbol{X}_1|\tilde{\boldsymbol{X}}}(t)|_{t=-\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{\beta}}}\big\} q\big(h_0, \big(- \tilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{\beta}}_0, \tilde{\boldsymbol{X}}^T\big)^T\big) \\
&\qquad + f_{\boldsymbol{X}_1|\tilde{\boldsymbol{X}}}\big(- \tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{\beta}}_0\big)\big\{\frac{\partial}{\partial t} q\big(h_0, \big(t, \tilde{\boldsymbol{X}}^T\big)^T\big)\Big|_{t=-\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{\beta}}_0}\big\}\big]\big) \\
=&E_{\tilde{\boldsymbol{X}}}\big(- \tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^T\big[\big\{\frac{\partial}{\partial t} f_{\boldsymbol{X}_1|\tilde{\boldsymbol{X}}}(t)|_{t=-\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{\beta}}}\big\} \times \Big(F_{T^*\left(0\right)|\boldsymbol{X}} - F_{T^*\left(1\right)|\boldsymbol{X}}\Big)\big[m_0 \mid \boldsymbol{X} = \big(- \tilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{\beta}}_0, \tilde{\boldsymbol{X}}^T\big)^T\big] \\
&\quad + f_{\boldsymbol{X}_1|\tilde{\boldsymbol{X}}}\big(- \tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{\beta}}_0\big)\big[\frac{\partial}{\partial t}\Big(F_{T^*\left(0\right)|\boldsymbol{X}} - F_{T^*\left(1\right)|\boldsymbol{X}}\Big)\big\{m_0 \mid \boldsymbol{X} = \big(t, \tilde{\boldsymbol{X}}^T\big)^T\big\}\Big|_{t=-\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{\beta}}_0}\big]\big])
\end{aligned}
$$

This $(p-1)\times(p-1)$ matrix is assumed negative definite in Condition (C4).

Furthermore, by (9), the consistency of $\widetilde{\boldsymbol{\beta}}_n$, the property of Kaplan-Meier estimator and the fact that $\Gamma\big(\widetilde{\boldsymbol{\beta}}_0, h_0\big) = 0$, it is straightforward to show $\Gamma\big(\widetilde{\boldsymbol{\beta}}_0, \widehat{h}\big) = o_p\big(n^{-1/2+\gamma_0}\big)$ for an arbitrarily small $\gamma_0 > 0$. So $||\Gamma\big(\widetilde{\beta}_0, \widehat{h}\big)|| = O_p\big(n^{-1/3}\big)$. Hence, Condition (B3) of DVK is also satisfied. Summarizing the above, we have proved the lemma.□.

Proof of Theorem 1 . We first introduce some new notation. Define

$$
\begin{aligned}
B\big(\boldsymbol{\beta}, h\big) &= Pg\big(\cdot, \boldsymbol{\beta}, h\big) - Pg\big(\cdot, \boldsymbol{\beta}_0, h\big), \\
B_n\big(\boldsymbol{\beta}, h\big) &= \mathbb{P}_n g\big(\cdot, \boldsymbol{\beta}, h\big) - \mathbb{P}_n g\big(\cdot, \boldsymbol{\beta}_0, h\big).
\end{aligned}
$$

And, for any $\delta > 0$, define a local upper bound function:

$$
M_\delta\big(\,\cdot\,\big) \geq \sup_{\|\widetilde{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}_0\| \leq \delta} \big|g\big(\cdot, \boldsymbol{\beta}, h_0\big) - g\big(\cdot, \boldsymbol{\beta}_0, h_0\big)\big|, \tag{10}
$$

[1] In the case where $\beta_{01} = \beta_1 = -1$, the gradient is the same formula except changing $-\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{\beta}}$ to $\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{\beta}}$ in every $q(h, \cdot)$; the Hessian matrix of $Pg\big(\cdot, \boldsymbol{\beta}, h\big)$ with respect to $\tilde{\boldsymbol{\beta}}$ evaluated at $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}_0, h = h_0$ is:

$$
\begin{aligned}
&\Lambda^*(\tilde{\boldsymbol{\beta}}, h)|_{\tilde{\boldsymbol{\beta}}=\tilde{\boldsymbol{\beta}}_0, h=h_0} \\
=&E_{\tilde{\boldsymbol{X}}}\big\{ - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^T\big(\big\{\frac{\partial}{\partial t} f_{\boldsymbol{X}_1|\tilde{\boldsymbol{X}}}(t)|_{t=\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{\beta}}}\big\} \times \Big(F_{T^*\left(1\right)|\boldsymbol{X}} - F_{T^*\left(0\right)|\boldsymbol{X}}\Big)\big[m_0 \mid \boldsymbol{X} = \big(\tilde{\boldsymbol{X}}^T\widetilde{\boldsymbol{\beta}}_0, \tilde{\boldsymbol{X}}^T\big)^T\big] \\
&\quad + f_{\boldsymbol{X}_1|\tilde{\boldsymbol{X}}}\big(- \tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{\beta}}_0\big) \times \big[\frac{\partial}{\partial t}\Big(F_{T^*\left(1\right)|\boldsymbol{X}} - F_{T^*\left(0\right)|\boldsymbol{X}}\Big)\big\{m_0 \mid \boldsymbol{X} = \big(t, \tilde{\boldsymbol{X}}^T\big)^T\big\}\Big|_{t=\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{\beta}}_0}\big]\big)\big\}.
\end{aligned}
$$

which dominates the class of functions $\mathcal{M}_\delta = \big\{g\big(\cdot,\boldsymbol{\beta},h_0\big) - g\big(\cdot,\boldsymbol{\beta}_0,h_0\big) : \|\widetilde{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}_0\| \le \delta, \beta_{01} = \beta_1\big\}$.

The proof involves verifying Conditions (C1)-(C10) of Theorem 3 in DVK. We first note that (C1) is satisfied by the proof for Lemma 1. Condition (C2) holds naturally in our case with $E = \mathbb{R}^{p-1}$. Condition (C6) is trivially satisfied under our regularity conditions. Condition (C8) is also satisfied since we constructed $\hat{\boldsymbol{\beta}}$ as the maximizer of the random function $\mathbb{P}_n g\big(\cdot,\boldsymbol{\beta},\hat{h}\big)$. We will verify the other conditions.

Condition (C3) can be verified using techniques similar as those for proving (B2) in proof of Lemma 1. More specifically, (C3) holds when the following two conditions are satisfied: there exist a function l and a positive constant $\delta_0 < \eta/2$ such that for all $\delta_2, \delta_3 < \delta_0$, and for an arbitrarily small $\gamma_0 > 0$,

$$n^{2/3} l\big(n^{-1/3}\delta_2, n^{-1/2+\gamma_0}\delta_3\big) = o(\sqrt{n}), \tag{11}$$

and

$$\begin{aligned} & E\Big[\sup_{\substack{\|\widetilde{\boldsymbol{\beta}}-\widetilde{\boldsymbol{\beta}}_0\| \le n^{-1/3}\delta_2,\ \beta_{01}=\beta_1, \\ d_{\mathcal{H}}\left(h,h_0\right)\le n^{-1/2+\gamma_0}\delta_3}} \Big| B_n(\boldsymbol{\beta},h) - B(\boldsymbol{\beta},h) - B_n(\boldsymbol{\beta},h_0) + B(\boldsymbol{\beta},h_0)\Big|\Big] \\ \le\ & n^{-1/2} l\big(n^{-1/3}\delta_2, n^{-1/2+\gamma_0}\delta_3\big). \end{aligned} \tag{12}$$

To verify the above conditions, let us consider the class:

$$\begin{aligned} \mathcal{F}_{\delta_2,\delta_3} = \big\{ & g\big(\cdot,\boldsymbol{\beta},h\big) - g\big(\cdot,\boldsymbol{\beta},h_0\big) - g\big(\cdot,\boldsymbol{\beta}_0,h\big) + g\big(\cdot,\boldsymbol{\beta}_0,h_0\big) : \\ & \|\widetilde{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}_0\| \le n^{-1/3}\delta_2,\ \beta_{01} = \beta_1, d_{\mathcal{H}}\big(h,h_0\big) \le n^{-1/2+\gamma_0}\delta_3\big\}, \end{aligned}$$

where $0 < \delta_0 < \eta/2$ and $0 < \delta_2, \delta_3 < \delta_0$.

It is straightforward to deduce that the random function $|g\big(\cdot,\boldsymbol{\beta},h\big) - g\big(\cdot,\boldsymbol{\beta},h_0\big)|$ is upper bounded by a constant $4\eta^{-2}n^{-1/2+\gamma_0}\delta_3$ (e.g., see (3)); as is the case with $|g\big(\cdot,\boldsymbol{\beta}_0,h\big) - g\big(\cdot,\boldsymbol{\beta}_0,h_0\big)|$. Therefore, an envelope function of the class $\mathcal{F}_{\delta_2,\delta_3}$ is:

$$F_{\delta_2,\delta_3}\big(Z\big) := 8\eta^{-2}n^{-1/2+\gamma_0}\delta_3. \tag{13}$$

It is straightforward to prove that $\mathcal{F}_{\delta_2,\delta_3}$ is a VC-class. Then $\mathcal{F}_{\delta_2,\delta_3}$ has finite uniform entropy integral. By Van Der Vaart and Wellner (1996), page 244, the left hand side of (12) is upper bounded by $\sqrt{E\big[F^2_{\delta_2,\delta_3}(Z)\big]}/\sqrt{n}$ up to a multiplicative constant. This, coupled with (13), inspired us to set $l\big(n^{-1/3}\delta_2, n^{-1/2+\gamma_0}\delta_3\big) = K_2 n^{-1/2+\gamma_0}\delta_3$ for a constant $K_2 > 8$. Then $\sqrt{E\big[F^2_{\delta_2,\delta_3}(Z)\big]} \le l\big(n^{-1/3}\delta_2, n^{-1/2+\gamma_0}\delta_3\big)$. Then, for all positive parameters $\delta_2, \delta_3 < \delta_0$, both (11) and (12) hold, implying that (C3) in DVK holds.

Condition (C4) concerns the function $M_\delta(\cdot)$ defined in (10). Observe that the envelope function $M_{\delta,\delta_1'}$ in (5) for the class $\mathcal{F}_{\delta,\delta_1'}$ does not depend on nuisance parameters and is also an envelope function for $\mathcal{M}_\delta$. So we let

$$M_\delta(\cdot) := 2\eta^{-1} I\Big[\frac{\delta}{\|\beta_0\|} > \frac{\big|\boldsymbol{X}^T\beta_0\big|}{\|\boldsymbol{X}\|\|\beta_0\|}\Big].$$

For $r_n = n^{1/3}$, we have

$$n^{-1}r_n^4 E\big[M^2_{\frac{K}{r_n}}\big] = n^{-1}r_n^4 O\big(Kr_n^{-1}\big) = O(1),$$

due to the assumption on the density of the angular component of $\boldsymbol{X}$. Hence, the first part of (C4) holds. $\forall\ \rho > 0$, $I\big\{r_n^2 M_{\frac{K}{r_n}}(\cdot) > \rho n\big\}$ is zero for all n sufficiently large. Hence,

$$n^{-1}r_n^4 E\big[M^2_{\frac{K}{r_n}} I\big\{r_n^2 M_{\frac{K}{r_n}} > \rho n\big\}\big] \le n^{-1}r_n^4 \frac{4}{\eta^2} P\big\{M_{\frac{K}{r_n}}(\cdot) > \rho r_n\big\} = o(1).$$

So the second part of (C4) also holds. By similar calculation, we can show (C5) holds.

By DVK's remark 3(iv), (C7) is fulfilled by defining the random function $W_n : \mathbb{R}^{(p-1)} \to \mathbb{R}$ as

$$W_n(t) = \big\langle \Gamma(\boldsymbol{\beta}_0, \hat{h}), t\big\rangle, \tag{14}$$

and define the deterministic bilinear function $V : \mathbb{R}^{(p-1)} \times \mathbb{R}^{(p-1)} \to \mathbb{R}$ as

$$V\big(t,t\big) = \frac{1}{2} t^T \Lambda\big(\beta_0, h_0\big) t, \tag{15}$$

where Γ and Λ are defined in the proof for Lemma 1.

Next we check (C9). In the proof of Theorem 1, we have shown $\Gamma\big(\widetilde{\boldsymbol{\beta}}_0, \hat{h}\big) = o_p\big(n^{-1/2+\gamma_0}\big)$

for an arbitrarily small $\gamma_0 > 0$. Hence, for an arbitrary compact subset $\mathcal{K}$ of $\mathbb{R}^{p-1}$,

$$n^{1/3} \sup_{t\in\mathcal{K}, t\neq 0} \|W_n(t)\|t\|^{-1}\| = o_{P^*}(1).$$

Since $\Gamma(\widetilde{\boldsymbol{\beta}}_0, h_0) = 0$,

$$\begin{aligned}
& r_n^2 B_n\big(\boldsymbol{\beta}_0 + \big(0, r_n^{-1}t^T\big)^T, h_0\big) \\
=& n^{2/3}\big[\mathbb{P}_n g\big(\cdot, \boldsymbol{\beta}_0 + \big(0, n^{-1/3}t^T\big)^T, h_0\big) - \mathbb{P}_n g\big(\cdot, \boldsymbol{\beta}_0, h_0\big)\big] \\
=& n^{2/3}\Big[\mathbb{P}_n g\big(\cdot, \boldsymbol{\beta}_0 + \big(0, n^{-1/3}t^T\big)^T, h_0\big) - \mathbb{P}_n g\big(\cdot, \boldsymbol{\beta}_0, h_0\big) \\
& - Pg\big(\cdot, \boldsymbol{\beta}_0 + \big(0, n^{-1/3}t^T\big)^T, h_0\big) + Pg\big(\cdot, \boldsymbol{\beta}_0, h_0\big)\Big] \\
& + \frac{1}{2}t^T\Lambda\big(\tilde{\boldsymbol{\beta}}, h\big)|_{\tilde{\boldsymbol{\beta}}=\widetilde{\boldsymbol{\beta}}_0, h=h_0} t + o(1),
\end{aligned} \tag{16}$$

where the exact form of Λ depends on whether $\beta_{01} = 1$ or $\beta_{01} = -1$, as discussed in the proof of Lemma 1. Therefore, the deterministic continuous function $\Psi(t)$ required in (C9) can be set to

$$\Psi(t) = \frac{1}{2}t^T\Lambda\big(\widetilde{\boldsymbol{\beta}}_0, h_0\big)t. \tag{17}$$

It remains to derive the limiting process of the first term on the right side of (16). Let $\mathbb{G}_n = \sqrt{n}\big(\mathbb{P}_n - P\big)$. Then the first term on the right side of (16) simplifies to

$$\begin{aligned}
& n^{2/3}\big(\mathbb{P}_n g\big[\cdot, \big\{\beta_{01}, \big(\tilde{\boldsymbol{\beta}}_0 + n^{-1/3}t\big)^T\big\}^T, h_0\big] \\
& \qquad - \mathbb{P}_n g\big(\cdot, \boldsymbol{\beta}_0, h_0\big) - Pg\big[\cdot, \big\{\beta_{01}, \big(\tilde{\boldsymbol{\beta}}_0 + n^{-1/3}t\big)^T\big\}^T, h_0\big] + Pg\big(\cdot, \boldsymbol{\beta}_0, h_0\big)\big) \\
=& n^{1/6}\big\{\mathbb{G}_n\big[g\big(\cdot, \big\{\beta_{01}, \big(\tilde{\boldsymbol{\beta}}_0 + n^{-1/3}t\big)^T\big\}^T, h_0\big) - g\big(\cdot, \boldsymbol{\beta}_0, h_0\big)\big]\big\}.
\end{aligned} \tag{18}$$

Note that this is the same as the centered process in the parametric case (see page 292 in Van Der Vaart and Wellner (1996)). The covariance function of the limiting process is

$$K(s_1, s_2) = \lim_{n\to\infty} n^{1/3} P\big\{g\big(\cdot, \big\{\beta_{01}, \big(\widetilde{\boldsymbol{\beta}}_0 + n^{-1/3}s_1\big)^T\big\}^T, h_0\big) g\big(\cdot, \big\{\beta_{01}, \big(\widetilde{\boldsymbol{\beta}}_0 + n^{-1/3}s_2\big)^T\big\}^T, h_0\big)\big\}. \tag{19}$$

Finally, the mean-zero process (18) converges weakly to a Gaussian process with covariance

function $K(s_1, s_2)$, which we denote by $\mathbb{W}(t)$.

To prove (C10), we can use similar techniques used in proving (B2) to prove $\mathcal{M}_\delta$ has finite uniform entropy integral, which is an alternative to the original condition in (C10) in terms of the bracketing integral (see (3.2.8) in Van Der Vaart and Wellner (1996)).

Summarizing the above, we have proved Conditions (C1)-(C10) of DVK hold. By Theorem 3 of DVK, $n^{1/3}\big(\widetilde{\boldsymbol{\beta}}_n - \widetilde{\boldsymbol{\beta}}_0\big)$ converges to the unique maximizer of the process $t \mapsto \Psi(t) + \mathbb{W}(t)$. $\square$

Proof of Lemma 1. Given an IDR $\boldsymbol{d} = \boldsymbol{d}_{\boldsymbol{\xi}} \in \mathcal{D}$, define

$$m_{\boldsymbol{\xi}}\left(b, X, D, Y, \Delta, G\right) = \frac{\widetilde{R}(\boldsymbol{d})\rho_\tau(Y-b)}{\widetilde{w}_{\boldsymbol{d}}},$$

where $\widetilde{R}(\boldsymbol{d}) = \Delta I\big(D_1 = d_{1,\boldsymbol{\beta}}(\boldsymbol{X})\big)\big\{I(Y \le s) + I(Y > s)I\{D_2 = d_{2,\boldsymbol{\zeta}}(H_2)\}\big\}$. Then $\widehat{Q}_\tau(\boldsymbol{\xi}, \widehat{G}_C)$ minimizes $M_n(b, \widehat{G}_C) = \mathbb{P}_n\{m_{\boldsymbol{\xi}}(b, X, D, Y, \Delta, \widehat{G}_C)\}$, where $\mathbb{P}_n$ stands for the empirical average.

Let $M(b, G_C) = E\{m_{\boldsymbol{\xi}}(b, X, D, Y, \Delta, G_C)\}$. we will first show that

$$M(b, G_C) = E\big(\rho_\tau(T^*(\boldsymbol{d}) - b\big). \tag{20}$$

Importantly, (20) implies that the marginal quantile $Q_\tau\{T^*(\boldsymbol{d}_{\boldsymbol{\xi}})\}$ minimizes $M(b, G_C)$. Note that (20) follows from two observations. First,

$$\begin{aligned}
& E\Big(\frac{\Delta I\big(D_1 = d_1(\boldsymbol{X}_1)\big)I(Y \le s)\rho_\tau(Y-b)}{\widetilde{w}_{\boldsymbol{d}}} \mid O^*\big(\boldsymbol{d}\big)\Big) \\
= \; & E\Big(\frac{I(T \le C)I\big(D_1 = d_1(\boldsymbol{X}_1)\big)}{\pi_{D_1}(\boldsymbol{X}_1)G_C(Y)} I(Y \le s)\rho_\tau(Y-b) \mid O^*\big(\boldsymbol{d}\big)\Big) \\
= \; & E\Big(\frac{I(\dot{T}\left(d_1, \emptyset\right) \le C)}{G_C(\dot{T}\left(d_1, \emptyset\right))} I\big(\dot{T}\left(d_1, \emptyset\right) \le s\big)\rho_\tau(\dot{T}\left(d_1, \emptyset\right) - b) \mid O^*\big(\boldsymbol{d}\big)\Big) \\
= \; & E\Big(I(\dot{T}\left(d_1, \emptyset\right) \le s)\rho_\tau(\dot{T}\left(d_1, \emptyset\right) - b) \mid O^*\big(\boldsymbol{d}\big)\Big)
\end{aligned}$$

.

Second,

$$
\begin{aligned}
& E\Big(\frac{\Delta I\big\{D_1 = d_1(\boldsymbol{X}_1)\big\} I\big\{D_2 = d_2(\boldsymbol{X}_1, D_1, \boldsymbol{X}_2)\big\} I(Y > s)\rho_\tau(Y - b)}{\widetilde{w}_{\boldsymbol{d}}} \mid O^*\big(\boldsymbol{d}\big)\Big) \\
= \; & E\Big(\frac{I(T \le C) I\big\{D_1 = d_1(\boldsymbol{X}_1)\big\} I\big\{D_2 = d_2(\boldsymbol{X}_1, D_1, \boldsymbol{X}_2)\big\}}{\pi_{D_1}(\boldsymbol{X}_1)\pi_{D_2}(\boldsymbol{X}_1, D_1, \boldsymbol{X}_2) G_C(Y)} I(Y > s)\rho_\tau(Y - b) \mid O^*\big(\boldsymbol{d}\big)\Big) \\
= \; & E\Big(\frac{I(\dot{T}\,(d_1, d_2) \le C)}{G_C(\dot{T}\,(d_1, d_2))} I\big(\dot{T}\,(d_1, \emptyset) > s\big)\rho_\tau\big(\dot{T}\,(d_1, d_2) - b\big) \mid O^*\big(\boldsymbol{d}\big)\Big) \\
= \; & E\Big(I\big(\dot{T}\,(d_1, \emptyset) > s\big)\rho_\tau\big(\dot{T}\,(d_1, d_2) - b\big) \mid O^*\big(\boldsymbol{d}\big)\Big).
\end{aligned}
$$

Hence, (20) holds.

To prove that $\widehat{Q}_\tau(\boldsymbol{\xi}, \widehat{G}_C) \to Q_\tau\left\{T^*\left(d_{\boldsymbol{\xi}}\right)\right\}$ in probability, we will verify that the following three conditions are satisfied.

(i) $\sup_{b<L}\big|M_n\big(b, \widehat{G}_C\big) - M\big(b, G_C\big)\big| \xrightarrow{P} 0$,

(ii) $\forall\ \epsilon > 0$, $\inf\big\{M\big(b; G_C\big) : \big|b - Q_\tau\left\{T^*\left(d_{\boldsymbol{\xi}}\right)\right\}\big| \ge \epsilon\big\} > M\big(Q_\tau\left\{T^*\left(d_{\boldsymbol{\xi}}\right)\right\}, G_C\big)$,

(iii) $M_n\big(\widehat{Q}_\tau\big(\boldsymbol{\xi}, \widehat{G}_C\big), \widehat{G}_C\big) \le M_n\big(Q_\tau\left\{T^*\left(d_{\boldsymbol{\xi}}\right)\right\}, \widehat{G}_C\big) + o_P\big(1\big)$.

To prove (i), note that

$$
M_n\big(b, \widehat{G}_C\big) - M\big(b, G_C\big) = \big[M_n\big(b, \widehat{G}_C\big) - M_n\big(b, G_C\big)\big] + \big[M_n\big(b, G_C\big) - M\big(b, G_C\big)\big]. \tag{21}
$$

By the preservation theorems in §2.10.2 of Van Der Vaart and Wellner (1996), the class of functions $\mathcal{J} = \big\{m_b\big(X, A, Y, \Delta; G_C\big) : b \in (0, L), \boldsymbol{\xi} \in \{-1, 1\} \times \widetilde{\mathbb{B}} \times \{-1, 1\} \times \widetilde{\mathbb{Z}}\big\}$ is *Donsker*. Hence,

$$
\sup_{b<L}\big|M_n\big(b, G_C\big) - M\big(b, G_C\big)\big| \xrightarrow{a.s.} 0. \tag{22}
$$

By Csörgő and Horváth (1983), $\forall\ \gamma_0 > 0$, $\sup_{t<L}\big|\widehat{G}_C\big(t\big) - G_C\big(t\big)\big| = o\big(n^{-1/2+\gamma_0}\big)$ almost surely. Therefore, for $\forall\ \gamma_0 > 0$,

$$
\begin{aligned}
& \sup_{b<L}\big|\boldsymbol{M}_n\big(b, \widehat{G}_C\big) - \boldsymbol{M}_n\big(b, G_C\big)\big| \le 2\sup_{t<L}\big|\widehat{G}_C\big(t\big) - G_C\big(t\big)\big|\Big|n^{-1}\sum_{i=1}^{n}\eta^{-2}\rho_\tau(Y_i - b)\Big| \\
= & o\big(n^{-1/2+\gamma_0}\big) O_p(1).
\end{aligned} \tag{23}
$$

(23) together with (22) and (21) imply (i). Next, (ii) holds because $M\big(b, G_C\big)$ is continuous and uniquely minimized at $b = Q_\tau\{T^*(d_{\boldsymbol{\xi}})\}$. Lastly, (iii) is implied by the fact that the criterion function $M_n\big(\cdot, \widehat{G}_C\big)$ is minimized at $\widehat{Q}_\tau\big(\boldsymbol{\xi}, \widehat{G}_C\big)$.

We will next show the above properties lead to the consistency of $\widehat{Q}_\tau\big(\boldsymbol{\xi}, \widehat{G}_C\big)$. From (ii), for any small $\epsilon > 0$, $\exists\ \vartheta > 0$ such that for b such that $\big|b - Q_\tau\{T^*(d_{\boldsymbol{\xi}})\}\big| \geq \epsilon$, we have $\boldsymbol{M}\big(b; G_C\big) > \boldsymbol{M}\big(Q_\tau\{T^*(d_{\boldsymbol{\xi}})\}; G_C\big) + \vartheta$. Thus

$$
\begin{aligned}
& P\Big(|\widehat{Q}_\tau\big(\boldsymbol{\xi}, \widehat{G}_C\big) - Q_\tau\{T^*(d_{\boldsymbol{\xi}})\}| \geq \epsilon\Big) \\
\leq\ & P\Big(\boldsymbol{M}\big(\widehat{Q}_\tau\big(\boldsymbol{\xi}, \widehat{G}_C\big); G_C\big) > \boldsymbol{M}\big(Q_\tau\{T^*(d_{\boldsymbol{\xi}})\}; G_C\big) + \vartheta\Big) \\
\leq\ & P\Bigg(\boldsymbol{M}\big(\widehat{Q}_\tau\big(\boldsymbol{\xi}, \widehat{G}_C\big); G_C\big) - \boldsymbol{M}_n\big(\widehat{Q}_\tau\big(\boldsymbol{\beta}, \widehat{G}_C\big); \widehat{G}_C\big) > \\
& \qquad \boldsymbol{M}\big(Q_\tau\{T^*(d_{\boldsymbol{\xi}})\}; G_C\big) - \boldsymbol{M}_n\big(Q_\tau\{T^*(d_{\boldsymbol{\xi}})\}, \widehat{G}_C\big) + \vartheta/2\Bigg)(1 + o(1)) \\
\leq\ & P\Big(\sup_{b<L}\big|\boldsymbol{M}\big(b, G_C\big) - \boldsymbol{M}_n\big(b, \widehat{G}_C\big)\big| > \vartheta/4\Big)(1 + o(1)) \to 0,
\end{aligned}
$$

as $n \to \infty$, where the second inequality follows from property (iii) and the last step follows from property (i). Thus $\widehat{Q}_\tau\big(\boldsymbol{\xi}, \widehat{G}_C\big) \xrightarrow{P} Q_\tau\{T^*(d_{\boldsymbol{\xi}})\}$. □

2 Additional numerical results

Example S1 (comparing quantile criterion and mean criterion) In this example, we compare the performance of the new method with the method of Simoneau et al. (2020) which considers a mean-optimal criterion. the method of Simoneau et al. (2020) is implemented using the function `DWSurv` in the R package `DTRreg` (see Wallace et al. (2020)). The goal is to illustrate that each has its own advantage depending on the specific target. Here, we generate the random data from a heteroscedastic regression model $\log(T) = (X_1 + X_2) + A(-X_1 + X_2)^3 + (0.5 + A * (0.5X_1 + 0.5X_2))\epsilon$ where X_1 and X_2 are independent random variables with Uniform$(0, 1)$ distribution, $\epsilon \sim N(0, 0.5)$ is independent of $\boldsymbol{X} = (X_1, X_2)^T$, and $A \sim$ Bernoulli(0.5) is independent of $(\boldsymbol{X}, \epsilon)$. The censoring variable C is generated from the Uniform$[0, 13]$ distribution. The censoring rate of the data is about 24%. Table 1 summarizes the simulation results for $n = 1000$ based on 300 simulation runs. We consider

the quantile-optimal IDR for $\tau = 0.25$ and $\tau = 0.5$, respectively. We evaluate the estimated 0.25 quantile ($\widehat{Q}_{0.25}$), 0.5 quantile ($\widehat{Q}_{0.5}$) and mean ($\widehat{\mu}_{\text{mean}}$) using a large independent sample of size 10^7. The results suggest that the different optimal IDRs perform best with respect to their own targets.

	$\widehat{Q}_{0.25}$	$\widehat{Q}_{0.5}$	$\widehat{\mu}_{\text{mean}}$
New ($\tau = 0.25$)	**2.220**	3.045	3.298
New ($\tau = 0.5$)	1.883	**3.290**	3.427
Simoneau et al.	2.080	3.102	**3.453**

Table 1: Results for Example S1

Example S2 (comparing the model-free method with model-based methods) In general, different methods that target different optimality criteria lead to different optimal IDRs. In this example, we consider a model where different criteria yield the same optimal rule. In this setting, the estimated parameters corresponding to the optimal IDR are comparable. We focus on comparing the proposed model-free method with two model-based methods in two scenarios: (i) the outcome regression model is correctly specified; (ii) the outcome regression model is mis-specified. Specifically, we compare with the semiparametric method of Simoneau et al. (2020) and a model-based approach using censored quantile regression by Peng and Huang (2008). The approach of Simoneau et al. (2020) requires a model for the blip function (or knowledge of the treatment-covariate interaction effect). The approach of Peng and Huang (2008) targets the conditional quantile and requires to specify the full model for the outcome regression. In this example, the approach of Simoneau et al. (2020) fits a linear blip function, while the approach of the approach of Peng and Huang (2008) fits a censored quantile regression with the linear main effect and linear covariates and interaction effect. The method of Peng and Huang (2008) is implemented using the function `crq` in the R package `quantreg` (see Koenker et al. (2020)).

We first consider the setting of correct model misspecification. the log-transformed survival time is generated from $\log(T) = (-0.2+0.5X_1-0.5X_2)+1.2A(0.2-0.6X_1-0.1X_2)+\epsilon$, where X_1 and X_2 are independent random variables with Uniform$(0,1)$ distribution, $\epsilon \sim N(0, 0.5)$ is independent of $\boldsymbol{X} = (X_1, X_2)^T$, and $A \sim$ Bernoulli(0.5) is independent of $(\boldsymbol{X}, \epsilon)$. The censoring variable C is generated from the Uniform$[0, 5]$ distribution. The censoring

rate of the data is about 29%. One can show that for the above location-scale model with i.i.d. random errors, the quantile-optimal treatment regime (no matter the quantile level τ of interest) is given by $\mathrm{I}(0.2 - 0.6X_1 - 0.1X_2 > 0)$. This would also be the same optimal treatment regime for the expected mean optimal criterion.

Table 2 summarizes the bias and standard deviations for estimating the parameters indexing the optimal IDR, where the coefficient of X_1 is normalized to have absolute value 1, for $n = 1000$ based on 300 simulation runs. It also reports the mis-classificiation rate (MR), the average percentage of times the estimated optimal IDR does not match the theoretically optimal IDR. All three methods perform well in this scenario with the model-based methods having slightly smaller standard errors.

	Intercept	X_2	MR
New ($\tau = 0.5$)	0.006	-0.027	0.009
	(0.102)	(0.150)	
New ($\tau = 0.25$)	0.009	-0.036	0.011
	(0.095)	(0.151)	
Simoneau et al.	-0.014	0.029	0.007
	(0.058)	(0.098)	
Peng& Huang ($\tau = 0.5$)	-0.001	0.001	0.0005
	(0.064)	(0.114)	
Peng& Huang ($\tau = 0.25$)	-0.001	0.004	0.001
	(0.073)	(0.116)	

Table 2: Results for Example S2 (correct model specification)

Next, we compare the performance of the three methods under model misspecification. The random data are generated the same as above except that $\log(T) = (-0.2 + 0.5X_1 - 0.5X_2) + 1.2A(0.2 - 0.1X_2 - 0.6X_1)^3 + \epsilon$. The censoring rate of the data is about 28%. Here the interaction effect is nonlinear but a monotone transformation of the linear index. The theoretically optimal IDR remains the same for all three methods. However, both Simoneau et al. (2020) and Peng and Huang (2008) fitted using linear models and hence are misspecified. Table 3 summarizes the results. We observe that the proposed new method still accurately estimates the optimal IDR while the other two methods both have considerable bias with inflated mis-classification rates.

Example S3 (the case with more covariates) We generate the survival time from the

	Intercept	x1	x2	MR
New ($\tau = 0.5$)	-0.012	0.000	-0.009	0.016
	(0.148)	(0.000)	(0.160)	
New ($\tau = 0.25$)	-0.005	0.000	-0.026	0.018
	(0.152)	(0.000)	(0.172)	
Simoneau et al.	-0.132	0.180	-0.173	0.177
	(2.932)	(0.573)	(5.372)	
Peng& Huang ($\tau = 0.5$)	-0.541	0.240	0.104	0.250
	(8.721)	(0.651)	(9.101)	
Peng& Huang ($\tau = 0.25$)	-0.523	0.253	1.682	0.503
	(6.069)	(0.666)	(32.166)	

Table 3: Results for Example S2 (model mis-specification)

model $\log(T) = \boldsymbol{X}^T\boldsymbol{\alpha} + A(\boldsymbol{X}^T\boldsymbol{\beta}) + \epsilon$, where $\boldsymbol{X} = (1, X_1, \ldots, X_p)^T$ with the components being independent Uniform$(0, 1)$ random variables, $\epsilon \sim N(0, 0.5)$ is independent of $\boldsymbol{X}$, and $A \sim$ Bernoulli(0.5) is independent of $(\boldsymbol{X}, \epsilon)$. The censoring variable C is generated from the model $\log(C) = -0.7 + 0.8X_1 - 0.2X_2 + 0.6X_3 + 0.8X_4 + \epsilon$. We consider $p = 10$ and 20.

- For $p = 10$, $\boldsymbol{\alpha} = (-0.2, 0.5, -0.5, 0.3, -0.4, 0.5, -0.2, 0.4, -0.1, 0, 0)$ and $\boldsymbol{\beta} = (0.5, -0.6, -0.1, 0.1, 0.5, 0, -0.6, 0.1, -0.6, 0.3, -0.3)$.

- For the $p = 20$, $\boldsymbol{\alpha} = (0.5, 0.6, -0.3, 0, 0, 0, 0, 0.4, 0, \ldots, 0)$ and $\boldsymbol{\beta} = (0.5, 1, 0.2, 0.3, -0.3, -0.2, -0.4, 0.5, -0.5, 0.4, 0.6, 0.5, -0.6, -0.5, 0.4 - 0.4, 0, 0.3, -0.2, 0.2, 0.3)$.

The censoring rate of the data is about 0.26 for 10 dimension; and about 0.28 for 20 dimension. Table 4 summarizes the average mis-classification rate (the average percentage of times the estimated optimal IDR does not match the theoretically optimal IDR) for dimensional 10 and 20. We observe that the genetic algorithm still yields sufficiently accurate estimation of the optimal IDR although the computation becomes substantially more intensive.

n	$p = 10$	$p = 20$
500	0.066	0.028
1000	0.052	0.034
2000	0.043	0.046

Table 4: Average mis-classification rates for Example S3

References

Csörgő, S. and Horváth, L. (1983). The rate of strong uniform consistency for the product-limit estimator. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 62(3):411–426.

Delsol, L. and Van Keilegom, I. (2020). Semiparametric m-estimation with non-smooth criterion functions. *Annals of the Institute of Statistical Mathematics*, 72(2):577–605.

Koenker, R. et al. (2020). *quantreg: Quantile Regression*. R package version 5.86.

Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association*, 103(482):637–649.

Simoneau, G., Moodie, E. E., Nijjar, J. S., Platt, R. W., Investigators, S. E. R. A. I. C., et al. (2020). Estimating optimal dynamic treatment regimes with survival outcomes. *Journal of the American Statistical Association*, 115:1531–1539.

Van Der Vaart, A. W. and Wellner, J. A. (1996). Weak convergence. In *Weak Convergence and Empirical Processes*, pages 16–28. Springer.

Wallace, M., Moodie, E. E. M., Stephens, D. A., Simoneau, G., and Schulz, J. (2020). *DTRreg: DTR Estimation and Inference via G-Estimation, Dynamic WOLS, Q-Learning, and Dynamic Weighted Survival Modeling (DWSurv)*. R package version 1.7.